

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**

**THIS PAGE BLANK (USPTO)**

**O'Bryen, Barbara**

---

**Fr m:** Ramirez, Delia  
**Sent:** Tuesday, December 30, 2003 3:41 PM  
**T :** STIC-Biotech/ChemLib  
**Subject:** case 09/983,025

Hi,

I would like to request the following searches:

1. a standard search of seq id 2 in the protein databases (commercial and interference)
2. an oligo search of seq id 2 in the protein databases (commercial).

Thank you,

---

Delia M. Ramirez, Ph.D.  
Patent Examiner  
Recombinant Enzymes-Art Unit 1652  
USPTO  
1911 S. Clark Street, Crystal Mall 1, 10D06, Mail room 10D01  
Arlington, VA 22202  
(703) 306-0288  
delia.ramirez@uspto.gov

**THIS PAGE BLANK (USPTO)**

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2004, 16:02:37 ; Search time 26 Seconds  
(without alignments)  
2914.567 Million cell updates/sec

Title: US-09-983-025a-2  
Perfect score: 9856  
Sequence: 1 MMCLIKLRISLAIAGWALC.....AADCDDECTCRDPKABENQ 1791

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database: Issued Patents AA:  
1: /cgn2\_6/ptodata/1/1aa/5A COMB pep.\*  
2: /cgn2\_6/ptodata/1/1aa/5B COMB pep.\*  
3: /cgn2\_6/ptodata/1/1aa/6A COMB pep.\*  
4: /cgn2\_6/ptodata/1/1aa/6B COMB pep.\*  
5: /cgn2\_6/ptodata/1/1aa/PCUS COMB pep.\*  
6: /cgn2\_6/ptodata/1/1aa/backfile1.pdp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	287.5	2.9	1847	6	Patent No. 5256642-10
2	287.5	2.9	1847	6	Patent No. 5472939
3	287.5	2.9	2039	6	Patent No. 5256642-2
4	287.5	2.9	2039	6	Patent No. 5472939
5	287.5	2.9	1466	6	Patent No. 5256642-6
6	283	2.9	1466	6	Patent No. 5472939-6
7	283	2.9	1537	6	Patent No. 5256642-5
8	283	2.9	1537	6	Patent No. 5472939
9	249	2.5	830	5	Patent No. 5472939-5
10	243.5	2.5	830	5	Patent No. 5472939-4
11	236	2.4	611	4	Patent No. 5472939-2
12	236	2.4	611	4	Patent No. 5472939-3
13	236	2.4	611	4	Patent No. 5472939-4
14	235.5	2.4	574	4	Patent No. 5378464-32
15	235	2.4	574	4	Patent No. 5378464-3
16	230.5	2.3	610	1	Patent No. 5378464-2
17	230.5	2.3	610	1	Patent No. 5378464-3
18	230.5	2.3	610	4	Patent No. 5378464-3
19	230.5	2.3	610	4	Patent No. 5378464-3
20	225	2.3	1394	6	Patent No. 5217870-2
21	211	2.1	1833	5	Patent No. 517197-30
22	211	2.1	1833	5	Patent No. 517197-30
23	209.5	2.1	1833	5	Patent No. 517197-30
24	209.5	2.1	324	1	Patent No. 517197-30
25	206.5	2.1	324	1	Patent No. 517197-30
26	191	1.9	2556	3	Patent No. 517197-30
27	191	1.9	2556	3	Patent No. 517197-30

28	187.5	1.9	1253	3	US-08-479-722B-4	Sequence 4, Appl
29	187	1.9	484	2	US-08-252-493C-9	Sequence 9, Appl
30	187	1.9	484	3	US-09-276-197-9	Sequence 9, Appl
31	183	1.9	2471	1	US-08-185-432-16	Sequence 16, Appl
32	183	1.9	2471	1	US-08-083-590A-19	Sequence 19, Appl
33	183	1.9	2471	1	US-08-532-384-19	Sequence 19, Appl
34	183	1.9	2471	4	US-08-899-232-1	Sequence 5, Appl
35	180.5	1.8	3075	2	US-08-460-309-5	Sequence 5, Appl
36	180.5	1.8	3075	2	US-08-125-077-5	Sequence 5, Appl
37	180	1.8	1251	5	PCT-US95-02251-3	Sequence 3, Appl
38	180	1.8	1252	1	US-08-199-780-3	Sequence 3, Appl
39	180	1.8	1252	1	US-08-316-650-3	Sequence 3, Appl
40	179	1.8	810	2	US-08-820-170A-34	Sequence 34, Appl
41	179	1.8	810	3	US-09-055-699-34	Sequence 34, Appl
42	179	1.8	810	3	US-09-273-565-34	Sequence 34, Appl
43	179	1.8	810	4	US-09-565-538-34	Sequence 34, Appl
44	179	1.8	810	4	US-09-661-468-34	Sequence 34, Appl
45	179	1.8	810	4	US-09-976-165-34	Sequence 34, Appl

#### ALIGNMENTS

```

RESULT 1
5256642-10
Patent No. 5256642
APPLICANT: PEARON, DOUGLAS T.; KLICKSTEIN, LLOYD B.; WONG,
WINNIE W.; CARSON, GERRARD R.; CONCINO, MICHAEL F.; IT, STEPHEN
H.; MARIDES, SAVVAS; MARSH, HENRY C. JR.
TITLE OF INVENTION: COMPOSITIONS OF SOLUBLE COMPLEMENT
RECEPTOR 1 (CR1) AND A THROMBOTIC AGENT, AND THE METHODS OF
USE THEREOF
NUMBER OF SEQUENCES: 30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/588,128
FILING DATE: 24-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 412,745
FILING DATE: 26-SEP-1989
APPLICATION NUMBER: 332,865
FILING DATE: 03-APR-1989
APPLICATION NUMBER: 176,532
FILING DATE: 01-APR-1988
SEQ ID NO:10:
LENGTH: 1847
5256642-10
Query Match 2.9%; Score 287.5; DB 6; Length 1847;
Best Local Similarity 19.1%; Pred. No. 4.7e-15;
Matches 349; Conservative 178; Mismatches 548; Indels 751; Gaps 105;
454 LVLTASEPVNT-EMVDF-----RDE-----KYPRLVLOGFEPPELISLPPL-- 498
39 LALPVAMGQCNABEMLFARPTNLTDFEFPFGYLVNVEGRPGVSGRFSIILCKNSVMT 98
499 -----CGQTYCDN-----VELISYNGVPLRGKVI 525
99 GAKDRCKRCKRNPDPVGMVHVIKIQGSOIKKSTGYRILIGSSATCIISGDTVI 158
526 RYGVNVCID--DEGLNPIVSEQIRLOHEALNBAF--SHYNISWQSLVHVNSTLRHR 580
159 WDEETPCIDILPGGLPPTIR-----NODFISTNENRHY-----GS 194
581 VLVNCEPSKIGND-----HCDPECEHPLTYGDG--DCRLQGRCYSNRRDGL 627
195 VVTVRCNPGSGGKRVFELVGEPSIYCTSNDDQ--VGISGPAPQCIIPNCTPPNVNGI 252
628 CHVECNMM--LNFDDGDC-----C-----DPOVADVAKTGFDPDPSPRAY 666
253 LVSDNLSLFLNVEVERCQPVFMKPRRVKCOALNKWPELPSGSRVQOPPDVLAH- 311
667 MSVKELEALQLNSTPST--LNIYFASVREDLAGAATW-----PMDKDAVT----- 710

```

[illegible]

```

OY      1441 QKEILLTSSSHMONTN-CLPUDCGVDPDSLVYANAFSCSE--TKFLKRSISGCV--- 14355
Db      1305 ---CVLWGMRSILMNNVSVPVCEHIFCPNP-PAILNGRHGTGPSGDI PYGKEISTYCDPHD 1360
OY      1496 ---PAKLOQSFPWLTCEd-----GLMSLPEVEYCKL-----ECDAPIILANLILPHCLQ 1543
Db      1361 RGMTPNLIGEST-IRCHSDPHNGWMSAPARCLSVRAGCKTKPEQFPFSPITP--IN 1417
OY      1544 D-NHDVGTICKYCKPQGYVAESAEGKVRNKLKIOCLEGGIWE--QSCIPVVCPEPP 1600
Db      1418 DPEFPVGTSLNYECRPGYF-----GKMFISCIENLWMSVEDNCRKSCGPPPE 1467
OY      1601 VFEGM-----VECTNGFSL---DQCVLNQNQREKXLPILCTKSGLWTOE 1642
Db      1468 PFNGVNHINDTQFGSTVNNVSCNEGFRILGSPSTTCLVSGNNV-----TWDKK 1515
OY      1643 PKLCENLQGEPPPEPPEISN-----VEYKCEQGYG----- 1672
Db      1516 APIGEIIT--SEEPPTISNGDFPENNRTSFHNGTVVYIQCHTGDGDLPELVGBRSIYC 1573
OY      1673 -----IGAVCSP-----LCVIPP-SDFVMLPENIT---ADTEHMMEP---VKV 1709
Db      1574 TSKDDQGVWMSPPPRCISTNKTCAFEVENAIRVGNRSPFSLTEIIRFCQPGFVWGS 1633
OY      1710 QSIYCGRQWHPDPVLVHCIOSECP 1735
Db      1634 HTVQCQTNGRW--GPKLPHCSRVCOP 1657

RESULT 2
5472939-10
Patent No. 5472939
Applicant: FEARON, DOUGLAS T.; KLICKSTEIN, LLOYD B.; WONG,
WINNIE W.; CARSON, GERALD R.; CONCINO, MICHAEL F.; IP, STEPHEN
H.; MAKRIDES, SAVVAS; MARSH, HENRY C. JR.
TITLE OF INVENTION: METHOD OF TREATING COMPLEMENT
MEDIATED DISORDERS
NUMBER OF SEQUENCES: 30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/138,825
FILING DATE: 19-OCT-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 588,128
FILING DATE: 24-SEP-1990
APPLICATION NUMBER: 412,745
FILING DATE: 26-SEP-1989
APPLICATION NUMBER: 332,865
FILING DATE: 03-APR-1989
APPLICATION NUMBER: 176,532
FILING DATE: 01-APR-1986
SEQ ID NO:10:
5472939-10
LENGTH: 2006

Query Match      2.9%; Score 287.5; DB 6; Length 1847;
Best Local Similarity 19.1%; Pred. No. 4.7e-15;
Matches 349; Conservative 178; Mismatches 548; Indels 751; Gaps 105;

OY      454 LVLTASFEPPVNT-EWVPF-----RDE-----KYPLEVLQGFEPEDILSPLOPL- 498
Db      39 LALPVAWQCNAPMLPFPARPTNLNDEFEPFIGTYLVNVECRPGYSGRPFISICLKNSVWT 98
OY      499 -----CGQTVDN-----VELISQYNGWPLRGEVY 525
Db      99 GAKDRCRRKSKGRNPDPVNGVNHVIKGIQFGSQIKYSCTKGYRILIGSSSATCIIISGDTVI 158
OY      526 RYGVVNICD--DEGLNPVSEEQIRLOHEALNEAF-----SRYNISMQLSVHQVHNSLRLHR 580
Db      159 WDNETPICDRIKPLPLPIT-----NDFISTNKNRNFY-----GS 194
OY      581 VLVNCEBSKIGND-----HCDPECEHPLTGYDGG--DCLQGRCSYNNRRDGL 627

```

```

Db 195 VTYRCPNGSGGRVFEVGPSEIYCTSNDDQ--VGIWSPAPQCIIPNKCTPPNENGI 252
Qy 628 CHEVCNNM--LNDEDDGDC-----C-----DPOVADVKTCPDPSPKRAY 666
Db 253 LVSNRBLFSLNEVERRCQVFWKGRBRKVCQALNKMBELPSCSRVCPDPDLHA- 311
Qy 667 MSYKELKALQJNSTHF--LNIYFASVREDLAQAATW-----PMDKAVT----- 710
Db 312 -----ERTQRDKNFSPQGEVYFSCPGYDLGAASMRCTPGQDMSPAPTEVNSCD 364
Qy 711 -----HGGIYLSAAYGMFCHTDTMHEVGHVIG--LHVFG-----VSERE 752
Db 365 DFMGQLNGRYLFPVNLQJAKVDFVCEGFOJLKGSSASVCLAGMESLMNSVAVCEQI 424
Qy 753 SCNDPCKEYVPS-METG-----DLCAD----- 773
Db 425 FC--PSPVYFNGHHTGKPLEVFPFGKAVNYTCDPHPDRGTSFPLIGSTIRCTSDQGN 482
Qy 774 ---TAPTKSEL--CREPE-----PTSDTCGTFRP--GAFETVMS 808
Db 483 GWMSAPARCGILGHCAQPHFLFAKLTQTNASDFIGISLKYECREYVGRFES-- 538
Qy 809 YTDNCTDNF--TPNOYARNHCTLDLVYQWTSRKPTPIPIPMV----- 852
Db 539 ---ITCLDNLVWSSPKVCK-----RKSKCTPPDPVNGMVHVTIDQVGSRIIN 583
Qy 853 -----IGQTKSLTI-----HM--LPPI-----SGVYVDRASGSLGACTEDGER 891
Db 584 YSCITGRRLIGHSSABECLISGNAAMWSTKPICORIPGCLPTIANGPFI-----STNR 637
Qy 892 QYVHTAS--SRVCDSSGYWTPBEAVGPDDV-----DQCEPSLQAMS--PEVHLYHNM 942
Db 638 ENFHYGVWYTRCNPGSGGRKVFELVGPSEIYCTSNDDQ-----VGIWSPAPQCIIPNK 992
Qy 943 TVPCPTGCSLELLFQHPVQADTLTMTWTSFPMSSQVLFPTTEILLE--NKSEVHIGPL 999
Db 693 CTPEPNE-----NGILVDSNLSFLNEVVERRCQPGFVWKGPR 731
Qy 1000 DTFCDIPLTILKIHVDGKVSQVYTPDERIGIDAALLISOPHSPLCSG--CRVAVYQVLRD 1058
Db 732 RVKQ-----ALNKMBELPSCSRVCP----- 754
Qy 1059 PPFASGLPVVHTSHRKFTDVE--VTPGMYOYQVLAENGEL--SEAS--PPLNHTIGA 1112
Db 755 -----PPDLAHERQORDKNFSPQGEVYF--CEBGYDLRGAASMRCTPGQDMSPA 805
Qy 1113 PYCG-----DGK--VSERIGEE--CDDGDLVSDGCS----- 1140
Db 806 PTCVKSQCDPFGQJLNGRVLFPVNLQJAKVDFVCEGFOJLKGSSASVCLAGMESLMN 865
Qy 1141 ---KVCE-----LE-----EGFVCVGPSEI--C 1158
Db 866 SSUPVCEQIIFCSPPIVNGHHTGKPLEVFPFGKAVNYTCDPHPDRGTSFPLIGESTIRC 925
Qy 1159 YM-YEGDGI-----CE-----PEERKTSIVDCGIYT-----PKG-- 1187
Db 926 TSDPQNGWSSAPARCGILGHCAQPHFLFAKLTQTNASDFIGISLKYECREYVGR 985
Qy 1188 -----LDQATRAYSHED--KKKC--PVSLVTEGPHSLI-----CTSYHPD 1225
Db 986 PFSITICLD--NLVWSSPKVCKRKSKCTPPDPVNGMVHVTIDQVGSRIINYSCTTGH-R 1041
Qy 1226 LPNHR-----PLTGWPRC-----VASENTQDRSSQPGSLKKEDEV 1263
Db 1042 LIGHSSABECLISGNTAMWSTKPICORIPGCLPTIANGPFISTNRENFHYGSV----- 1095
Qy 1264 WLKVCNRPGEARAI-----IFLTLDG-----LVGHEHQPTT--LYL 1301
Db 1096 -VTRCNLGSGRKRVFELVGPSEIYCTSNDDQVGIWSPAPQCIIPNKCTPPNENGI 1154
Qy 1302 TDVAGSNHSL-----GTYGLSCQ-----HNPLIINTYHQNVLFEH 1336
Db 1155 SD-----NRSJFSLNEVVDPRCQPGFVWKGPRRVCQALNKMBELPSCSRVCPDPDLH 1210

```

```

Qy 1337 --HTTSVLNFPSSPRVGISAVALRTSSRIGLASPNSCISEDEGONHOGOSCIHRPCKQ- 1393
Db 1211 GEHTPSHODNFS-----GOEVYSC---EPGYDLGAASLH--CTPQ 1249
Qy 1394 -----DSCPSLL--LDHADVNTCSIGGLMKCAITCQGFALQASSQGYIRPM 1440
Db 1250 DMSPEARCAVKSQCDDELQQLPHGRVLPPLNLQJG-AKVSFVCEGFRLLKSSVSH----- 1304
Qy 1441 QKEILLTSSGHWQDONS-CLPVDGVPDPSLVNYANFSCSEG--TKLKRCSISCV- 1485
Db 1305 ---CVLWGMBSLMWNSVPEVHEHIFCPNP-PALINGHRTGTPSGDI PYGKEISYTCDEHPD 1360
Qy 1496 ---PAKQGLSPMLTCLD-----GLWSLPREYCKL-----ECDAPIILNANLLPHCLQ 1543
Db 1361 RGMTEFLIGEST--IRCTSDPHNGWSSAPARCELSRAGHCKTPEQPPFASPTIP--IN 1417
Qy 1544 D-NHDVETICKYECPGYVAESAEGKVRNKLKIQCLEGGIWB--QGSCLPVVCEPPPP 1600
Db 1418 DFEFPVGTSLNTECRPGYF-----GKMFISCTLENLWMSVBDNCRKSCGPPPE 1467
Qy 1601 VFBSM-----YECTNGESL--DSQVLCNQERKLPILCTKEGLTQ 1642
Db 1468 PFNGMVHINTDQFGSTVWNSCNEGFRLLGSPSTTCLVSGNNV-----TWDK 1515
Qy 1643 FKLCENLQGCPPPESELS-----VEYKEQGY----- 1672
Db 1516 APIEIT--SCEPPTISNDFYSNNRTSFHNGTVYQCHTGDEGQJLPELVGERSTYC 1573
Qy 1673 ---IGAVCSP-----LCVIRP--SDPVMLEPNT--ADTLEHNMPE--VKV 1709
Db 1574 TSXQDQGVWSSPPRCISTNKCTAPEVENAIRVGNRSFSLTEILIRFCQPGFVWGS 1633
Qy 1710 QSIYCTGRQMHDPVULVHCLQSCPE 1735
Db 1634 HTVQOQTNGRW--GPKLPHCSRVCP 1657

RESULT 3
5256642-2
; Patent No. 5256642
; APPLICANT: FEARON, DOUGLAS T.; KLICKSTEIN, LLOYD B.; WONG,
; WINNIE W.; CARSON, GERALD R.; CONCINO, MICHAEL F.; ID, STEPHEN
; H.; MAXKIDS, SAVVAS; MARSH, HENRY C. JR.
; TITLE OF INVENTION: COMPOSITIONS OF SOLUBLE COMPLEMENT
; RECEPTOR 1 (CRI) AND A THROMBOTIC AGENT, AND THE METHODS OF
; USE THEREOF
; NUMBER OF SEQUENCES: 30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/586,128
; FILING DATE: 24-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 412,745
; FILING DATE: 26-SEP-1989
; APPLICATION NUMBER: 332,865
; FILING DATE: 03-APR-1989
; APPLICATION NUMBER: 176,532
; FILING DATE: 01-APR-1988
; SEQ ID NO:2:
5256642-2
LENGTH: 2039

Query Match 2.9%; Score 287.5; DB 6; Length 2039;
Beet Local Similarity 19.1%; Pred. No. 5.5e-15;
Matches 349; Conservative 178; Mismatches 548; Indels 751; Gaps 105;

Qy 454 LVLTASPEPVNT--EWVP-----RDE-----KYRLEVLQGEPEPEILSPLOPL-- 498
Db 34 LALPVANGQCNABEMLPFAPNTLDTFERPIGYLNYECPGSGRPFSTICLAKSVMT 93
Qy 499 ---CGQVCDN-----VELISQYNGVPLRGEKYI 525
Db 94 GAKDRCRRKSCRPPDPVNGMVHVIKIQFGSQIKYCTKGYRLIGSSATCISGDTVI 153

```

QY 526 RYQVNNICD---DEGLNPIVSEQRIQLQHEALNEAF---SRYNISQMSLVHQNHSILRHR 580  
 DB 154 WDNETPICDRI PCCLPPTIT---NGDFISTRENPHY-----GS 189  
 QY 581 VLVNCEPSKXIGND-----HCDPECEHPLTGYDGG---DCLRQRCYSWMRRDGL 627  
 DB 190 VVTVRCRPGSGGRKRVFELVGEPSIYCTISNDQ---VGIWSGAPQCIIIPKCTPPNVEGCI 247  
 QY 628 CHVECNMM---LNDDDGDC-----C-----DPQVADVKTCTDPSPRAY 666  
 DB 248 LVSNRSLFSLNEVEERCPVFVWKGBRRVCKQALNKMELBELPSCSRVCCPPDVLHA- 306  
 QY 667 MSVKELEALQINSTHF---LNIYFASSVREDLAGAATW-----PMDKAVT----- 710  
 DB 307 -----ERTRORDKDNFSPGQEVFYSCBPGYDLKGAASMRCTPGDMSPAAPCEVXSCD 359  
 QY 711 -----HLGGIYLSPAYGMPGHTDTMHEVGHVLG---LYHVFKG-----VSBRE 752  
 DB 360 DFMQQLNGRVLFPVNIQLAKVD FVCDGEGFLKSSASVCLAGMESLMSVPCVCEQI 419  
 QY 753 SCNDPCKEYPS-METG-----DLCAD----- 773  
 DB 420 FC--PSPPVIPNGHRTGKPLEVFPFGKAVNYTCDPHPDRTGSPDLIGESTRICTSDPGN 477  
 QY 774 ---TAPPKSEL---CREPE-----PTSDTCGTRRP---GAPFTYMS 808  
 DB 478 GWSPPAPRCGILGHCOAPDHFILFAKLTQTNASDPFGISLKECREYEGRPSS--- 533  
 QY 809 YTDNCTCNF---TPNOVARMHCYLDLVYQOMTESRKPTPIPPMV----- 852  
 DB 534 ---ITCLDNLWSSPKVCK-----RKSCKPPDPVGMVHVITDIOVGRIN 578  
 QY 853 -----IGQTKSLTI-----HW-LRPI-----SGVYDASGSLCACTEDGFR 891  
 DB 579 YSCTTGRLIGHSSAECILSGNAAMWSTKPICORIPGLPPTIANGPFI-----STNR 632  
 QY 892 QYVHTAS---SRVCDSSGYMTPPEAVGPPDV-----DQCEPSIQAMS-PEVHLHYMM 942  
 DB 633 ENFYSGVTVTRCMPGSGGRKRVFELVGEPSIYCTISNDQ---VGIWSGAPQCIIIPK 687  
 QY 943 TVPCPTGCSLELLFOHPVQADLTLLWTSFPMSSQVLPDTEILLE---NKESVHLGPL 999  
 DB 688 CTPEPNE-----NGILVSDNRSLFSLNEVEERCPGPFVWKGBR 726  
 QY 1000 DTECDIPLITGLAHYDGVSKVYTFDERIEIDALLTQSHSPICSG-CRPVXYQLARD 1058  
 DB 727 RVKCO-----ALNKMELBELPSCSRVCCP----- 749  
 QY 1059 PPFAAGLPVVVTHSHRKPTDVE-VTPGOMYQYVLAEAGEL-GEAS---PPLNHINGA 1112  
 DB 750 -----PPDVLHAERTORDKDNFSPGQEVFYS-CEGTYDLRGAASMRCTPGDMSPA 800  
 QY 1113 PYCG-----DGR-----VSERLGEE---CDDGDLVSGDGS----- 1140  
 DB 801 PTCVKSQCDPMQGLNGRVLFPVNLQGAKVDFCDEGFLKSSASVCLAGMESLNM 860  
 QY 1141 ---KYCE-----LE-----EGFNCVGEPSL-C 1158  
 DB 861 SSVAVCEBQIFCPBPVPIPNGHRTGKPLEVFPFGKAVNYTCDPHPDRTGSPDLIGESTIRC 920  
 QY 1159 YM-YEGDGI-----CE-----PEPKTSIYDCGIYT-----PKG-- 1187  
 DB 921 TSDPGNGWSSPPAPRCGILGHCOAPDHFILFAKLTQTNASDPFGISLKECREYEGR 980  
 QY 1188 -----LDQWATRAYSSHED---KKKC---PVSLVYGEPSHL-----CTSYHPD 1225  
 DB 981 PFSITCLD---NIVWSSPKDVCKRSKCTPPDPVGMVHVITDIOVGRINYSCTTGH-R 1036  
 QY 1226 LPNHR-----PLTGWFPCC-----VASNETODDRSQOPBSLKEDEV 1263  
 DB 1037 LIGHSSAECILSGNTHAMSTKPICORIPCGLPPTIANGDFISTRENPHYGSV----- 1090

QY 1264 WLKVCNRPGEARAI-----IFLTDD-----LVPEHQOPTYT---LYL 1301  
 DB 1091 -VTVRCNLGSRGRKRVFELVGEPSIYCTISNDQVIGSGAPQCIIIPKCTPPNVEGCI 1149  
 QY 1302 TVVRGNSHL-----GTYGLSCQ-----HNPILINVTTHQNVLFH 1336  
 DB 1150 SD-----NRSFLFSLNEVDFRCQPGFVWKGBRRVCKQALNKMELBELPSCSRVCCPPDVLH 1205  
 QY 1337 ---HTSVLNLNPPSPRVGISAVALRTSIRIGLASPNCSISDEGONHOGSCIRHPCGKO- 1393  
 DB 1206 GEHTPSHQDNFSP-----GOEYFSC---EPGYDLGAASLH---CTPGC 1244  
 QY 1394 -----DSCPSLL---LDHADVNYCTSIGFLMKCAITCORFALQASQGYRPM 1440  
 DB 1245 DWSPEARCAVKSDDDLGQLPHGRVLPNLQAG-AKVSFVCEGRLKSSVSH----- 1299  
 QY 1441 QKEIILTCSSGHHQDNVS-CLPYDCGVPDLSLVNYANFSCSEG-TKFLKRCISCV- 1495  
 DB 1300 ---CVLGMRSILMNSVPCVEHIFCPNP-PALINGRHTGTPSGDI PYGKEISYTCDPHPD 1355  
 QY 1496 ---PAKQGLSPMLTCLD---GLMSLPEYCKL-----ECDAPIILNANULLPHCLQ 1543  
 DB 1356 RGMTEFLIGEST-IRCTSDPHGNGWSSPPAPRCGLSVRAGHCKTBEQPPFASPTIP--IN 1412  
 QY 1544 D-NHDVGTICKYCKECPGYVAESAEGVRNKLKIOCLEGIME--OGSCIPVVCBPPPP 1600  
 DB 1413 DFERPVGTSLNBERPRTF-----GMPFISLLENLWMSVSDNCRKSCGPPPE 1462  
 QY 1601 VFBGM-----YECTNGPSL---DSQVLANQBEREKLPICTKEGLTQE 1642  
 DB 1463 PFNGMVHINTDQGSTVWNSCNEGFRILGSPYTCIVSGNNV-----TWDK 1510  
 QY 1643 FKJCENLQGECPPPSELS-----VEYKEQGYG----- 1672  
 DB 1511 APICELI--SCBPTTISNGDFISNNRTSPFNNGTVVYQCHTGDGQOLFELVGEPSIYC 1568  
 QY 1673 ---IGAVCSP-----LCVIPP-SDPVMLENIT---ADTLEHMMER---VKV 1709  
 DB 1569 TSKDQGVWSSPPPRCISTNKTCTAPEVBNAIRPGRSFFSLTEILIRPQCPGFVWGS 1628  
 QY 1710 QSIYCTGRQHPDPVVLHVCIOGCEP 1735  
 DB 1629 HTVQCOTNGRW--GPKLPHCSRVCCP 1652

RESULT 4  
 5472939-2  
 ; Patent No. 5472939  
 ; APPLICANT: PEARON, DOUGLAS T.; KLICKSTEIN, LLOYD B.; WONG,  
 ; WINNIE W.; CARSON, GERALD R.; CONCIRO, MICHAEL F.; IP, STEPHEN  
 ; H.; MAKRIDES, SAVVAS; MARSH, HENRY C. JR.  
 ; TITLE OF INVENTION: METHOD OF TREATING COMPLEMENT  
 ; MEDIATED DISORDERS  
 ; NUMBER OF SEQUENCES: 30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/138, 825  
 ; FILING DATE: 19-OCT-1993  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 588, 128  
 ; FILING DATE: 24-SEP-1990  
 ; APPLICATION NUMBER: 412, 745  
 ; FILING DATE: 26-SEP-1989  
 ; APPLICATION NUMBER: 332, 865  
 ; FILING DATE: 03-APR-1989  
 ; APPLICATION NUMBER: 176, 532  
 ; FILING DATE: 01-APR-1988  
 ; SEQ ID NO: 2:  
 ; LENGTH: 2039  
 5472939-2

Query Match 2.9%; Score 287.5; DB 6; Length 2039;  
 Best Local Similarity 19.1%; Pred. No. 5.5e-15;  
 Matches 349; Conservative 178; Mismatches 548; Indels 751; Gaps 105;

```

QY 454 LVLTASEPVNT-EMVPE-----RDE-----KYPRLEVLQGEPEBELSLPQPLU-- 498
D 34 LALPVAQGCNAPRMLPAPRTNLTDEEFPIGYLNEBCKRGVSGRPFSLCKAKSVMT 93
QY 499 -CGQTYCDN-----VELISQNGWPLRGEKVI 525
D 94 GAKDRCKRCKSCRNPPDPVNGMVAHVKIQGSGQIKYSCCKRGYRLIGSSASALCISGDTVI 153
QY 526 RYQVNNICD--DEGLNPVSEBQIRLQHEALNEAF--SRYNIMQSLVHVNHTSLRHR 580
D 154 WDNETPICDRIPCGPPTIT-----NGDFISTNRNFHY-----GS 189
QY 581 VLVNCEPSKIGND-----HCDPECEHPLTGYDG--DCRLQGRCSYNNRRDGL 627
D 190 VVTYRCNPGSGGKRVFELVGEPSIYCTSNDDQ--VGIMSGAPPCIIIPNKTPEVNGCI 247
QY 628 CHVECANM--LNDDDGDC-----C-----DPOADVKTCTFDPSPRAY 666
D 248 LVSDNRSLFSLNEVVEFCQPVFMKGRVYKCOLNKMELPSCSRVCCPPDVLHA- 306
QY 667 MSVKELMALQTNSTHF--LNIYFASVREDLAGAATW-----PMDKAVT----- 710
D 307 -----ERTORDXDNFSPGOEVFYSCPGYDLKGAASMRCTPGQDWSPPAFTCEVXSCD 359
QY 711 -HLGIVLSPAYGMPGHTMTIHEVHVLG--LYHVFKG-----VSERE 752
D 360 DFMQGLNGRLVFPVNIQAKVDFVCDGEGQLKSSASYCVLAGMESLNNSSVPEVQI 419
QY 753 SCNDPCKEYVPS-METG-----DLCAD----- 773
D 420 FC--PSPPVJPNHNGHTKPLEVFPFGKAVNYTCDPHPRGTSFDLIGSTIRCTSDPQN 477
QY 774 -TAPRPSKL--CREPR-----PNSDTCGTRFP--GAFTYMS 808
D 478 GWSPPARCGILGHCAPDHFLPAKLTQTNASDFPGLTSLKCECPREYGRPS----- 533
QY 809 YTDNCTDNF--TPNOYARMHCYLDLVYQOMTESRKTPIPIPMV----- 852
D 534 -ITCLDNLWSSPKOVCK-----RKSCTPRPVAQMVHVTIDIOVGRIN 578
QY 853 -IGQNKSLTI--HW--LPI--SGVYDRASSGLCAGTEDGTFR 891
D 579 YSCTTGRLIGHSSAECLISGNAAMWSTKPIICORIPGILPPTIANGDFI-----STNR 632
QY 892 QYVITAS--SRVODSSGYWTEBAVGPBV--DQCEBPLQAMS--PEVHLHYMM 942
D 633 ENFHYGSVVTYRCNPGSGGKRVFELVGEPSIYCTSNDDQ--VGIMSGAPCIIIPNK 687
QY 943 TVPCPTGCSLELLFQHPVQADTLTLMWTSFPMSSQVLPDTEILLE--NKESVHLGFL 999
D 688 CTRPNVE-----NGILVSDNLSFLSLNEVVEPRCQPGVVMKGFPR 726
QY 1000 DTFCDIPLTIKLHVDKVGKVTYTPDERLEIDAALLTQSPHSPLCSG--CRPVRYOVLRD 1058
D 727 RVKCO-----ALNKMWBELPSCSRVCP----- 749
QY 1059 PPRFASGLPVVYTHSHRKTIVE-VTPQMYOYQLAENGSL--GEAS--PLNHLHGA 1112
D 750 -PPVLAHERQORDKNFSPGOEVFY--CEBGYDLRKAASRCTPQDWSPPAA 800
QY 1113 PYCG--DOK--VSERLGEE--CDDGDLVSGDGS----- 1140
D 801 PTCGVKSCDDPMQGLNGRLVFPVNIQAKVDFVCDGEGQLKSSASYCVLAGMESLNN 860
QY 1141 -KVCE-----LE-----EGFNCVGEPSL-C 1158
D 861 SSVAVCEQIFCSPBPVLPNGRHTGKPLEVFPFGKAVNYTCDPHPRGRSFDLIGESTIRC 920
QY 1159 YM-VEGQI-----CE-----PEERKTSYVDCGIYT-----PKGY-- 1187
D 921 TSDPQNGWSSPAPRCGILGHCAPDHFLPAKLTQTNASDFPGLTSLKCECPREYGR 980

```

```

QY 1188 -LDOWATRAYSHED--KKKC--PVSLVTEGPHSLI-----CTSYHPD 1225
D 981 PFSITCD--NLWVSSPKVCKRCKSCTPRDDPVNGMVAHVTIDIOVGRINYSCTGH-R 1036
QY 1226 LPMNR-----PLTGMFPC-----VASENETODDREOEGBGLKXDEV 1263
D 1037 LIGHSABCLISGNTAMSTKPIICORIPGLPPTIANGPFISTRNENFHYGSV----- 1090
QY 1264 MLKVCNRPPEARAF-----IFLTGDC-----LVPEHQOQPTT--LYL 1301
D 1091 -VTRYCNLGRKRVFELVGEPSIYCTSNDDQVIGSGAPPCIIIPNKTPEVNGCI 1149
QY 1302 TDVRSNHSU-----GTYGLSCQ-----HNPLINVTTHQNVLFH 1336
D 1150 SD--NRSFLSLNEVVDPRCQPVFMKGRVYKCOLNKMELPSCSRVCCPPPIILH 1205
QY 1337 -HTTSVLNFPSSPRVIGISAVALTSTRIGLSAPNSCISDEGONHOGQSCIRPCKQ- 1393
D 1206 GEHTPSHQDNFSP-----GOEVFYSC--EPGYDLRGAASLH--CTPQG 1244
QY 1394 -DSCPSLL--LDHADVYVNTSIGPGLMKCAITCORGALPASSQGYLRPM 1440
D 1245 DMSPEAPRCVYKSCDDPLGQLPHERVLPVNLQGL-AKVSFVCDGEGRLKSSVSH----- 1299
QY 1441 QKEBILTCSSGHDONVS-CLPYDCGVDPDSLAVNYANFSCSEG-TKFLKRCISICVP-- 1495
D 1300 -CVLWGRMSLNNNSVPEVHIFCPRP-PALINGRHTGTPSGDIPYKXESTYCDHPD 1355
QY 1496 -PAKIQSLSPWLTLED--GLWSLPEVYCKL-----ECDAPPIIILNLLPHCLO 1543
D 1356 RGMTFNLIGEST-IRCTSDPHGNGWSSPAPRCBLSVRACHKTPQPPASPTIP--IN 1412
QY 1544 D-NHDVGTICKYCKPQGYVYAESAEGKVRNKLKIQCLEGIE--OGSCIPVCEPRP 1600
D 1413 DFEPPVGTSLNVERCPYF-----GKMFISICLENLWVSSVDNCRKRS CGPPE 1462
QY 1601 VFBSM-----YECTNGPSL--DSQCVLNCNOERKLPILCTREGMTQIE 1642
D 1463 PFMGMVHINTDQFGSTVYVNSCNBGRFLIGSPITGLVSGNNV-----TWDKK 1510
QY 1643 KLCENLQECPPPSSELSN-----VEYKEQGYG----- 1672
D 1511 APICEIT--SCBPPPTISNDFYSNNRTSPHNGTVVYQCHTGPDDGQLFELVGERSIYC 1568
QY 1673 -IGAVCSP-----LCVIRP--SDPVLMBEIT--ADTLEHMEP--VKV 1709
D 1569 TSXDDQVWSSPPRCISTNKTAPREVENAIRVGNRSFSLTEIIRFCQPGFVWGS 1628
QY 1710 QSIYCTGRQMHDPVLAHCIOGCEP 1735
D 1629 HTVQCCOTNGRM--GPKLPHCSRVCP 1652

RESULT 5
5256642-6
PATENT NO. 5256642
APPLICANT: PEARON, DOUGLAS T.; KLICKSTEIN, LLOYD B.; WONG,
WINNIE W.; CARSON, GERALD R.; CONCINO, MICHAEL F.; LE, STEPHEN
H.; MARIKIDS, SAVVAS; MARSH, HENRY C. JR.
TITLE OF INVENTION: COMPOSITIONS OF SOLUBLE COMPLEMENT
RECEPTOR 1 (CR1) AND A THROMBOLYTIC AGENT, AND THE METHODS OF
USE THEREOF
NUMBER OF SEQUENCES: 30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/588,128
FILING DATE: 24-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 412,745
FILING DATE: 26-SEP-1989
APPLICATION NUMBER: 332,865
FILING DATE: 03-APR-1989
APPLICATION NUMBER: 176,532
FILING DATE: 01-APR-1988

```

SEQ ID NO: 6;  
LENGTH: 1466  
5256642-6

Query Match 2.9%; Score 283; DB 6; Length 1466;  
Best Local Similarity 19.5%; Pred. No. 7.8e-15;  
Matches 262; Conservative 140; Mismatches 454; Indels 490; Gaps 71;

602 HPLGVGGGCRLOGRCVSNRRDGLC-HVECNMNLNDFDGDCCDPQVADVARTCTCPDP 660  
Db HRLGHSSAECITLGNAAHSTKPCICRIPC-----LPTTIA----- 121  
QY 83 HRLGHSSAECITLGNAAHSTKPCICRIPC-----LPTTIA----- 121  
Db 661 SPKRAVSVKELKALQLNSTHFLNIYFASVREDLAGAATWPKDAVTHLGGIVLSPA 720  
QY 122 -----NGDFTSTNRENF-----HYGSVV---T 140  
Db 721 YYGMPGHTDTHMEVGHVGLY-----HVFKVSERESCNDPCKETVPDMETDGLCA 772  
QY 141 YRCNPGSGGRKVFELVGPISYCTSNDDQVIMSGPAPQ--CIIPNKCTPPNVENGILVS 198  
Db 773 DTAFT-----PKSELCR--EPEPTSDTCGFTFPGAPFTNMYST 810  
QY 199 DNRSLFSLNEVEFRCPGFVPMKGRVVKCOALNKWPELPSCSRVCP--PPDVLAHERT 257  
QY 811 DNDCTDNETPNOVARMHCYLDLVYQQWTESRKPPIPIPPMVIGQTNKSLTIHMLPISG 870  
Db 258 ORD-KDNFSPGOEVFYSC-----EPG----- 277  
QY 871 VVYD-RASGSLCGACTEDGTRQYVHTASRRVCDSSGYTPEEAVGPPD-----VDQP 923  
Db 278 --YDLRGASAM--RCTPQGDWSPAPATCEVKSCDPFMQLNGRLVFNQLGAKVDFV 333  
QY 924 CEPGLQAMSPVYHVMNTVPCPTBEGSLELPHQVADTLTLMVTFMESQVLF 983  
Db 334 CDGEPQ-----LKSSA--SYCLAGMESLMSNVV--CEQIFCP 370  
QY 984 TEILLENKESVHLG-PLDTF-----CDIPLTIKLHVDKSGVRYTFDERIEIDA 1033  
Db 371 SPVLPNGR--HTGKPLEVFPFGKAVNYTCD-----PHPRGTS-----FD-LIGEST 415  
QY 1034 ALLTSQPI-----SPL-----CGGCRPRYVGLDPPRPSGLPVVTHSHRKPLDV 1079  
Db 416 IRTSDPQGNWSPAPRCGILGHCOAPDHFLEFKLXTQTNASDPFI----- 463  
QY 1080 EVTPGQWYOVLAEAG-----ELGEASP-----PLNHIGAPY--CGDG 1118  
Db 464 ----GTSUKYECRPEYRGRPSITCLDNLVWSSPDKVCKRSKTPRPVGMNAVITDI 519  
QY 1119 KYSERLGEBCDGDVLVSGDGSKYCELEBGFNCVGPESLCYMYEGDGICEPFRKTSIVD 1178  
Db 520 QVGSRIINVSCTTGHRLIGHSSAE-----CISGNTAHMSTKPCICOR-----IP 563  
QY 1179 CG-----IYPRGYLDOMATRAYSSHEDKKKCVSLVTEGPHSLICTSHPD--- 1225  
Db 564 GCLPPTIANGDFISTNRENFHYGSVYTYRCNLGSGRGVTFELVGP--SIYCTSNDDQVGI 622  
QY 1226 -----LPHN--RPLTGMFPCVASENETODDSEQESGLKXEDVWLKVCFNRPSEA 1275  
Db 623 WSGRAPQCIIPNKCTP-----PNVENGILVSDNR--LFLSLNEVEFRCPGFVPMKGR 674  
QY 1276 RAIFILTTDGLVPEHQOQPTVTLVTLVDRGSNHSIGTYGLSCQHNPLIINVTH---HQN 1332  
Db 675 R-----VKCOALNKWPELPSCSRV-----CQPRPELIGHMTPSHQD 712  
QY 1333 VLFHHTTVLANFSSPRVGISAVALKRTSRIGLASPSNCISEBQNGOGSCIH----- 1387  
Db 713 -----NFSF-----GOEVFYSC--EPGYDRLGASALHCTPRG 742  
QY 1388 -----RPGKQDSCSL--LDHADVNVCTSIGPBLMCAITCGGFLQLQASGQYIRPM 1440  
Db 743 DMSAPRCAVAKSCDPLGOLPHGRVLPPLNLQLD-AVYSFYVCDGFLKSGSVSH----- 797  
QY 1441 QKEILLTSSGHDQNV--CLPVDGVPDPPLVNYANFSCSEG--YKFLKRGSIQCV-- 1495

Db 798 ---CIVGMRSLMNNSVPEVCEHIFCPNP--PAIINGRHTGPSGDIPIYKSEISYTCDBHPD 853  
QY 1496 ---PAKLGUSPMLTCLD-----GIMSLPREVYCKL-----ECDAPIILNANLLPHPCLO 1543  
Db 854 RGMTEFNIGST--IRCTSDHNGWSSPAPRELVSRACHKTPGPPASFTTP--IN 910  
QY 1544 D-NHDVGTICKYECCKPQYVAESAEGKVRNKLKIQCLEGIME--QGSCTPVVCEPPPP 1600  
Db 911 DFERPVTSILNRECRPFY-----GKMSISGLELWVSSVEDNRRRSCGPPPE 960  
QY 1601 VFEGM-----YECTNGPSL-----DSQVLNQQREKPLILCTEGMLTQF 1642  
Db 961 PFNGMVHINTDQGSTVSNVSCNEGFLIGSPSTTCLVSGANNV-----TWDDK 1008  
QY 1643 FKUCENIQGECPPRPSLNS-----VEYKEQGYG----- 1672  
Db 1009 APICEIT--SCEPPTISNGDFYSNNRTSFHNGTAVYVYQCHTBPDEGLFELVGERSIYC 1066  
QY 1673 ---IGAVCSP-----LCVIPP--SDPVMLENIT-----ADTLEHMMEP---VKV 1709  
Db 1067 TSKDDQVGWSSPPRRCISTNKCTAPREVENAIRVGRSFFSLTEIRFRCOGFVWVS 1126  
QY 1710 QSIVCTGRQWHPDPVLVHCIOGCEP 1735  
Db 1127 HTVOCOTNGRW--GPKLPHCSRVCOP 1150

## RESULT 6

5472939-6

Patent No. 5472939  
APPLICANT: FEARON, DOUGLAS T.; KLICKSTEIN, LLOYD B.; WONG,  
MINNIE W.; CARSON, GERALD R.; CONCINO, MICHAEL F.; IP, STEPHEN  
H.; MAKRIDES, SAVVAS; MARSH, HENRY C., JR.  
TITLE OF INVENTION: METHOD OF TREATING COMPLEMENT  
MEDICATED DISORDERS  
NUMBER OF SEQUENCES: 30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/138, 825  
FILING DATE: 19-OCT-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 588, 128  
FILING DATE: 24-SEP-1990  
APPLICATION NUMBER: 412, 745  
FILING DATE: 26-SEP-1989  
APPLICATION NUMBER: 332, 865  
FILING DATE: 03-APR-1989  
APPLICATION NUMBER: 176, 532  
FILING DATE: 01-APR-1988  
SEQ ID NO: 6;  
LENGTH: 1466  
5472939-6

Query Match 2.9%; Score 283; DB 6; Length 1466;  
Best Local Similarity 19.5%; Pred. No. 7.8e-15;  
Matches 262; Conservative 140; Mismatches 454; Indels 490; Gaps 71;

602 HPLGVGGGCRLOGRCVSNRRDGLC-HVECNMNLNDFDGDCCDPQVADVARTCTCPDP 660  
Db 83 HRLGHSSAECITLGNAAHSTKPCICRIPC-----LPTTIA----- 121  
QY 661 SPKRAVSVKELKALQLNSTHFLNIYFASVREDLAGAATWPKDAVTHLGGIVLSPA 720  
Db 122 -----NGDFTSTNRENF-----HYGSVV---T 140  
QY 721 YYGMPGHTDTHMEVGHVGLY-----HVFKVSERESCNDPCKETVPDMETDGLCA 772  
Db 141 YRCNPGSGGRKVFELVGPISYCTSNDDQVIMSGPAPQ--CIIPNKCTPPNVENGILVS 198  
QY 773 DTAFT-----PKSELCR--EPEPTSDTCGFTFPGAPFTNMYST 810  
Db 199 DNRSLFSLNEVEFRCPGFVPMKGRVVKCOALNKWPELPSCSRVCP--PPDVLAHERT 257

```

QY 811 DNDCTDNFTPNQVARMHCYLDLVYQOWTESRKPTPIPIPMVIGQTNKSLTIHMLPPISG 870
D 258 QRD-KDNFSGQEVFVSC-----EPG----- 277
QY 871 VVVD-RASGSLCGACTEDGTFRQYVHTASRRVCDSSGYWTPBEAVGPPD-----VDOP 923
D 278 --YDLRGAASM--RCTPDGDMSPAAPTCCEVKSODDFGQGLNGRVLPVVLQGAQKDFV 333
QY 924 CEBSLQAMSEVHLVHMMMTVPCTEGCSLELLFQHPVQADTLTLWTSFMESSOVLFD 983
D 334 CDEBFO-----LKGSSA---SYCVLAGMESLWNSVVP--CEQIFCP 370
QY 984 TEILLEKESVHLG-PLDTF-----CDIPLTIKLAVDGKVSQVYVTFDERIDIDA 1033
D 371 SPVIFNGR--HTGKPLEVFPFGKAVNYTCO-----PHDRGTS-----FD-LIGEST 415
QY 1034 ALLTSOPH-----SPL-----CSGCRPVRYOVLRDPPFASGLPVVVTSHRKRTDV 1079
D 416 IRCTSDPDQNGWSSPAPRCGILGHCOAPDHPFLFAKLTQTNASDPFI----- 463
QY 1080 EVTPGQMYOYQVLAEGG-----ELGBAP-----PLNTHGAPY--CGDG 1118
D 464 ----GTSKTECREPTEYGRPSTICLDNLWSSPKDYCKRKSCTPPDPVNGVHVTDI 519
QY 1119 KVSERLGEBCDDGDLVSGDGCSKYCELEBGFNCVGEPSLCYMEGDCICEPFEKTSIVD 1178
D 520 QVSGRIYVSTCTGHRILGHSSAE-----CLSGNTAHMSTKPRICQ-----IP 563
QY 1179 CG-----IYTPGYLDQWATRAYSSHEDKKCPVSLVTGEPSLICTSYHPD-- 1225
D 564 CGHPRIANGDFISTNENFHYGSVVYRCVLGSRGKVFELVGEPP--SIYCTSDNDQVGI 622
QY 1226 -----LPHN-RPLTGMPCVASENETODRSBQPEGLKKEDEVMILKVCNRRGEA 1275
D 623 WSGPAPCIIIPNKCTP-----ENVENGILVSDNS---LESILNVVPRCOPGVWKGPR 674
QY 1276 RAIFPLITDGLVGEHQOFTVTLVLTVDVRSNHSLSGTGYSQOHNELIINVTH--HON 1332
D 675 R-----VKQALNKWEPBELPSCSRV-----CQPPBELLGHETTPHQD 712
QY 1333 VLFHTHTSVLNFSSPVGISAVALRTSSRIGLSAPNSCISEDEGQNHQSGSCIH---- 1387
D 713 -----NFSP-----GQEVFVSC--EPGYLDRGAASLHCTPRG 742
QY 1388 -----RPGKQDQSPSL--LDHADVYNCTSIGRGLMKCATTCQGRPALQASSQOYRPM 1440
D 743 DMSPEARCAVKSODDFLQGLPHGRVLPPLNLQLG-AKVSFVCDGFRLLKSSVSH---- 797
QY 1441 QKEILLTCSSGHWQDQNTS-CLPVDGVPDPSELVANYANFSCSEG--TKFLKCSISCPV-- 1495
D 798 ----CVLVGMSLWNSVVPVCEHIFCPNP--PAIINGRHTGTPSGDIPIYGEKELSTYCDHPD 853
QY 1496 ---PAKLGSLPWLTCLD-----GLWSLPEVYCKL-----ECDAPIILANMLLPRLQ 1543
D 854 RGMFTNFIGEST--IRCTSDPHGNGWSSPAPRCCLSVRAGCHCKPBEQFPFASPPIP--IN 910
QY 1544 D-NHNDVCTICKYBCKPGVYVAESAEGKVRNKLKIQCLEGGIWE--OGSCIPVVCSEPPP 1600
D 911 DFEPPVGLSLNTECRPGYF-----GKMFSISCLEMLWNSVSDNCRKSCGPPPE 960
QY 1601 VFEGM-----YECTNGPSL--DSQCVLNCQOERELKPICTKEGLWTOE 1642
D 961 PFNMVHAINITDQGTSTVYNSCNGFRLLIGSPSTTCLVSGNNV-----TYDKK 1008
QY 1643 FKLCENIAGBCPPPSSELSN-----VEYKCEQGYG----- 1672
D 1009 APICEII--SCPEPPTISNGDFYGNMRTSFFNGTVVTVYQCHTGQDGLFELVGERSIYC 1066
QY 1673 ----GAVGSP-----LCYIIP--SDPYMLPENIT--ADTLEHMMER--YKV 1709
D 1067 TSKDDQGVWSSPPRCISTNKCTAPAEVENAIRVGNRSFSLTEIIRFCQPGFVWGS 1126
QY 1710 QSIIVCTGRQWHPDPVLVHCIOGCEP 1735

```

```

D 1127 HTVQCCQNGRW--GPKLPHCSRVCOP 1150

RESULT 7
5256642-5
; Patent No. 5256642
; APPLICANT: FEARON, DOUGLAS T.; KLICKSTEIN, LLOYD B.; WONG,
; MINNIE W.; CARSON, GERALD R.; CONCINO, MICHAEL F.; IP, STEPHEN
; H.; MARBIDS, SAYAS; MARSH, HENRY C. JR.
; TITLE OF INVENTION: COMPOSITIONS OF SOLUBLE COMPLEMENT
; RECEPTOR 1 (CRI) AND A THROMBOLYTIC AGENT, AND THE METHODS OF
; USE THEREOF
; NUMBER OF SEQUENCES: 30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/588,128
; FILING DATE: 24-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 412,745
; FILING DATE: 26-SEP-1989
; APPLICATION NUMBER: 332,865
; FILING DATE: 03-APR-1988
; APPLICATION NUMBER: 176,532
; FILING DATE: 01-APR-1988
; SEQ ID NO:5:
; LENGTH: 1537
5256642-5

Query Match 2.9%; Score 283; DB 6; Length 1537;
Best Local Similarity 19.5%; Pred. No. 8.4e-15;
Matches 262; Conservative 140; Mismatches 454; Indels 490; Gaps 71;

QY 602 HPLTGDGDCRCLOGRCSYKNRDGLC-HYECNNMLNDFPDGDCPPQVADVKTCPDP 660
D 83 HRLIGHSAACILISGNAAMHSTKPRICQRIPCG-----LPPITL----- 121
QY 661 SPRAVYVSELKEALQLNSTHFLINITYFASVVEDLAGATWPDWDXAVTHLGIVLSPA 720
D 122 -----NGDFISTNENF-----HYGSV--T 140
QY 721 YGMPGHDTMIHEGVHGLY-----HYKGVSEKSCNDPCKETVPSMETDGLCA 772
D 141 YRCNPGSGRKVLELGEPSIYCTSDNDQVIGWGPAPQ--CIIPNKCTPPVENGILVS 198
QY 773 DTPAT-----PKSELGR--BEPTSDTGGTFRPPAPPTNWSYT 810
D 199 DMSLPSLNVVPRCOPGVWKGPRVKCOALNKWEPBELPSCSRVCOP--PPVLAHERT 257
QY 811 DNDCTDNFTPNQVARMHCYLDLVYQOWTESRKPTPIPIPMVIGQTNKSLTIHMLPPISG 870
D 258 QRD-KDNFSGQEVFVSC-----EPG----- 277
QY 871 VVVD-RASGSLCGACTEDGTFRQYVHTASRRVCDSSGYWTPBEAVGPPD-----VDOP 923
D 278 --YDLRGAASM--RCTPDGDMSPAAPTCCEVKSODDFGQGLNGRVLPVVLQGAQKDFV 333
QY 924 CEBSLQAMSEVHLVHMMMTVPCTEGCSLELLFQHPVQADTLTLWTSFMESSOVLFD 983
D 334 CDEBFO-----LKGSSA---SYCVLAGMESLWNSVVP--CEQIFCP 370
QY 984 TEILLEKESVHLG-PLDTF-----CDIPLTIKLAVDGKVSQVYVTFDERIDIDA 1033
D 371 SPVIFNGR--HTGKPLEVFPFGKAVNYTCO-----PHDRGTS-----FD-LIGEST 415
QY 1034 ALLTSOPH-----SPL-----CSGCRPVRYOVLRDPPFASGLPVVVTSHRKRTDV 1079
D 416 IRCTSDPDQNGWSSPAPRCGILGHCOAPDHPFLFAKLTQTNASDPFI----- 463
QY 1080 EVTPGQMYOYQVLAEGG-----ELGBAP-----PLNTHGAPY--CGDG 1118
D 464 ----GTSKTECREPTEYGRPSTICLDNLWSSPKDYCKRKSCTPPDPVNGVHVTDI 519
QY 1119 KVSERLGEBCDDGDLVSGDGCSKYCELEBGFNCVGEPSLCYMEGDCICEPFEKTSIVD 1178

```

```

Db 520 QVGRINVSCTTGRHLIGHSAB-----CILSGNTAHMSTKPIQCR-----IP 563
Qy 1179 CG-----YTPKGYLDQMATRAYSHEDKKKCPVSLVTEPHSLICTSYHPD--- 1225
Db 564 CGLPPTIANGDFISTNNENFYGSVVTYRCMLGSRGRVPELVGP-SIYCTSNDDQVGI 622
Qy 1226 -----LBNH-RPLTGWFPVCASBENOTDRESEDEGSLKKEDEWMLKVCENRGEA 1275
Db 623 WSGRAPCIIINCKCP-----PWNENGLVSDNRS---LFSLNEVEFERCQPGVFMKQPR 674
Qy 1276 RAIFILTTDLVGEHQOFTVTLVLDVRSNHSLSGTGLSCQNPILINVTN---HON 1332
Db 675 R-----VKCQALNKMBELPSCSRV-----CQPPPEILHGEHTPSHOD 712
Qy 1333 VLFHHTTSLVLFSSPRVIGISAVALTSSRIGLSAPNSCISBDEGQNHOGSCIH----- 1387
Db 713 -----NFSP-----GQEVFYSC-----EPGYDLRGAASLHCTPRG 742
Qy 1388 -----RPGKODSCPSLL--LDHADVNCISIGPLMKCAITCQGFALQASSGOYIRPM 1440
Db 743 DMSPPAPCAVKSDDFLGQLPHGRVLPPLNLQLG-AKVSFVCDGFRLLKSSVSH----- 797
Qy 1441 QKEILLTSSGHWQNVN-CLPVDGVPDPSLVNYANFSCSEG-TKFLKRCISICVP--- 1495
Db 798 ---CVLVGMRSILMNSVPEVCEHIFCPNP-PAILNGRHTGTPSGDIPYKKEISYTCDPHAPD 853
Qy 1496 ---PAKLOGLSRWLTCLSD-----GLMSLPVYVCU-----BEDAPILINMLPHLO 1543
Db 854 RGMFTNLIGEST-IRCTSDPHNGWSSPAPRCISLVAGHCKTPEQPPFASPTIP-IN 910
Qy 1544 D-NHVDGTICKCYCKRYGVAESAEGKRNKLLKIQCLEGGIWE--QSCSIPVVECPERP 1600
Db 911 DFEFVGSILMYECCRPGYF-----GKMFSISCIENLVMSVEVDCNRKSKGPPRE 960
Qy 1601 VFEEN-----YECTNGFSL-----DSQCVINCNQERKLPILCTKEGLWTOE 1642
Db 961 PFNGVNHINTDTQFGSTVYVNSCNEGFRLIGSPSTCLVSGNNV-----TWDXK 1008
Qy 1643 FKLCENLGECRPPSELSN-----VEKCEBGYV----- 1672
Db 1009 APICETI--SCEPPTISNGDFYSNNRTSFHNGTVVYTOCHGHPDEQLFELVGRSITYC 1066
Qy 1673 ---IGAVCSP-----LCVIPP-SDPYMLPENIT---ADTLEHMMEP---YKV 1709
Db 1067 TSKDDQVWMSPPRCISTKCTAPBEVEMNARVGNNSFSLTILINFRQPGVWVG 1126
Qy 1710 QSIYCTGRRQHPDPLVHICIQSCER 1735
Db 1127 HTVQCQTNGRW--GPKLPHCSRVCP 1150

```

## RESULT 8

```

5472939-5
Patent No. 5472939
Applicant: FEARON, DOUGLAS T.; KLICKSTEIN, LLOYD B.; WONG,
MINNIE W.; CARSON, GERALD R.; CONCINO, MICHAEL F.; IP, STEPHEN
H.; MAKIDES, SAVVAS; MARSH, HENRY C. JR.
Title of Invention: METHOD OF TREATING COMPLEMENT
MEDATED DISORDERS
Number of Sequences: 30
Current Application Data:
Application Number: US/08/138,825
Filing Date: 19-OCT-1993
Prior Application Data:
Application Number: 568,128
Filing Date: 24-SEP-1990
Application Number: 412,745
Filing Date: 26-SEP-1989
Application Number: 332,865
Filing Date: 03-APR-1989
Application Number: 176,532
Filing Date: 01-APR-1988

```

```

;SEQ ID NO:5:
5472939-5
LENGTH: 1537
Query Match 2.9%; Score 283; DB 6; Length 1537;
Best Local Similarity 19.5%; Pred. No. 8,4e-15;
Matches 262; Conservative 140; Mismatches 454; Indels 490; Gaps 71;
Qy 602 HPLTGYDGGDCRCYCSNNRRDGLC-HVECNMLMDPDDGDCDPOVADVKTCDPD 660
Db 83 HRLIGHSSABECILSGNAAMHSTKPIQCRIPC-----LPTTIA----- 121
Qy 661 SPKAYNVKELKEALQNSTHFLNIFYASSVREDLAGAATWPKDAVTHLGIVLSPA 720
Db 122 -----NGDFISTNNENF-----HYGSV---T 140
Qy 721 YYGMPGHTDWMIHVEGVILY-----HVEGVSEBSNDPCKEYPSMETGLCA 772
Db 141 YRCNPGSGRVPFELVEPSIYCTSNDDQVIGSGAPQ--CIIPNCTPPNENGLIVS 198
Qy 773 DTAPT-----PKSELGR-----EPPTSOTGFTFPGAFPTNYSYT 810
Db 199 DNRSLFSLNEVEFERCQPGVFMKQPRVKCOALNKMBELPSCSRVCP-PPVYLHERT 257
Qy 811 DNCTDNFTPNQVARMHCYLDLVYQWTESRKPPIPIPPWVIGQTNKSLITHMLPDISG 870
Db 258 ORD-KDNFSPQGEVIFYSC-----EPG----- 277
Qy 871 VVTP-RASGSLGACTEDGTFRQYVHTASRRVDSGTYTPBEAVGPPD-----VDQ 923
Db 278 --YDLRGAASW--RCTQGGMSPAPAPCEVKSDDFLGQLNGRVLPPLNLQLGAKVDFV 333
Qy 924 CEPSELQAMSPREVLHYHNMNTVPCPTEGCSLELTFQHPVQADTLTWTSPFMESSQVLF 983
Db 334 CDEGFO-----LKSSA---SYVLAMGBELMNSVVP-CEQLFCP 370
Qy 984 TEILLENKESVHLG-PLDTF-----CDIPILIKLVHGKVSQVYVTFDERIBIDA 1033
Db 371 SPVYIPNGR-HTKPLLEVPFPFGKAVNYTCD-----PHPDGRGS-----FD-LIGEST 415
Qy 1034 ALLTSQPH-----SPL-----GSGCRPVYQVLRDPPFASGLVVVYTHSHRKPTDV 1079
Db 416 IRCTSDPQNGWSSPAPRCGILGHCOAPDHPFLAKLTQTNASDPPI----- 463
Qy 1080 EVTPGQVQYOVLABAGS-----ELGEASP-----PNIHIGARY-GGDG 1118
Db 464 ---GTSIKYECREPEYVGRPSTICLDNLVMSFKDYCKRKSCTPDPVNGMHVTTDI 519
Qy 1119 KYSERLGEBCDDGLVSGDGSKYCELEBGFNCVGBEPLCYMYEGDGICEBPERKTSIVD 1178
Db 520 QVGRINVSCTTGRHLIGHSAB-----CILSGNTAHMSTKPIQCR-----IP 563
Qy 1179 CG-----YTPKGYLDQMATRAYSHEDKKKCPVSLVTEPHSLICTSYHPD--- 1225
Db 564 CGLPPTIANGDFISTNNENFYGSVVTYRCMLGSRGRVPELVGP-SIYCTSNDDQVGI 622
Qy 1226 -----LBNH-RPLTGWFPVCASBENOTDRESEDEGSLKKEDEWMLKVCENRGEA 1275
Db 623 WSGRAPCIIINCKCP-----PWNENGLVSDNRS---LFSLNEVEFERCQPGVFMKQPR 674
Qy 1276 RAIFILTTDLVGEHQOFTVTLVLDVRSNHSLSGTGLSCQNPILINVTN---HON 1332
Db 675 R-----VKCQALNKMBELPSCSRV-----CQPPPEILHGEHTPSHOD 712
Qy 1333 VLFHHTTSLVLFSSPRVIGISAVALTSSRIGLSAPNSCISBDEGQNHOGSCIH----- 1387
Db 713 -----NFSP-----GQEVFYSC-----EPGYDLRGAASLHCTPRG 742
Qy 1388 -----RPGKODSCPSLL--LDHADVNCISIGPLMKCAITCQGFALQASSGOYIRPM 1440
Db 743 DMSPPAPCAVKSDDFLGQLPHGRVLPPLNLQLG-AKVSFVCDGFRLLKSSVSH----- 797
Qy 1441 QKEILLTSSGHWQNVN-CLPVDGVPDPSLVNYANFSCSEG-TKFLKRCISICVP--- 1495

```



Db 798 ---CIVGMRLMNSVPCVCEHIFCPNP-PALINGRHTGTPSCDIPYCKEISYCDHPD 853  
QY 1496 ---PAKIQGSPMLTCLD---GLMSLPEVYCKL---ECDAPIILNANLLPRLCQ 1543  
Db 854 RGMFTNLIGST-IRCTSDPHGNWSSPAPRCLSVRAGCKTPEOPFPASPTIP--IN 910  
QY 1544 D-NHVDYTCIKYCKECPGYVAASAGKVRNKLKIQCLEGGIWE--OGSCIPVVCSEPP 1600  
Db 911 DFERPVGTSLNBYERPGYF-----GKMFISICLENLWMSVDNCRKSCGEPPE 960  
QY 1601 VFEGM-----YECTNGFSL---DSQVLANCQERKLPILCTKGLWTQ 1642  
Db 961 PFNGMVAINDTQGSTVYNSCNEGFRILGSPSTTCLVSGNNV-----TWDXK 1008  
QY 1643 FKICENLQGCPRPSELNS-----VEYKCEQGYG----- 1672  
Db 1009 APICEIT--SCEPPTISNDGFYSNNRTSPHNGTVTYQCHTGDBQLFELVGERSIYC 1066  
QY 1673 ---IGAVCSP-----LCVIPP-SDPVMLEPNIT---ADTLEHMMEP---VKV 1709  
Db 1067 TSKDDQGVWSSPPEPRCISTNKCTAPBEVNAIRPKNRSPSLTEIRFCQGFVWVGS 1126  
QY 1710 QSIYCTGRQWHPDPVLVHCIOGCEP 1735  
Db 1127 HTVQCQTNGRW--GPKLPKHSRVCOP 1150

RESULT 9  
US-08-110-158-4  
Sequence 4, Application US/08110158  
Patent No. 5605821  
GENERAL INFORMATION:  
APPLICANT: McEvey, Rodger P.  
APPLICANT: Pan, Junliang  
TITLE OF INVENTION: Expression Control Sequences of the  
TITLE OF INVENTION: P-Selection Gene  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Patrea L. Pabst  
STREET: 1100 Peachtree Street, Suite 2800  
CITY: Atlanta  
STATE: GA  
COUNTRY: USA  
ZIP: 30309-4530  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/110.158  
FILING DATE: 19930820  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/320,408  
FILING DATE: 08-MAR-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Pabst, Patrea L.  
REGISTRATION NUMBER: 31,284  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (404)-815-6508  
TELEFAX: (404)-815-6555  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 830 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-110-158-4  
Query Match 2.5%; Score 249; DB:1; Length 830;

Best Local Similarity 22.1%; Pred. No. 2.7e-12;  
Matches 102; Conservative 59; Mismatches 160; Indels 140; Gaps 25;  
QY 1382 QGSCIH-RPGKGDSCPSLLBDHADVNCISTPGLMKCAITQGRGALQASSQYIRPM 1440  
Db 191 GPCEVYRECGELELPOHVLNMCNSHPKGNFSFN---SQCSFHTDGVQVNG-----PS 240  
QY 1441 QKELILTCSGSHV-DONVSLCPVDCGVDPDSLNVYANFSCSEGTKFLKR---CSISC--- 1493  
Db 241 KLECL--ASGITNKPQCLLAQC--PLKIERGMITLHAKAFQHQSSCSFSCEEG 295  
QY 1494 ---VPPAKIQGSPMLTCLDGLMSLPEVYCK-LBC---DAP-----PIILNANLL 1537  
Db 296 FALVGEVVG-----CTASGVWTAAPAPCKAVQCHLEAPSGMTDCVHPLTAFV--- 345  
QY 1538 LPHCLDNHNDVGTICKYCKECPGYVAASAGKVRNKLKIQCLEGGIWE--GSCIPVVC 1595  
Db 346 ---YGSCKFECPQGY-----RVRG-LDMRCISGHWASAPLPCEAISC 386  
QY 1596 EPPPVVEGMYECT---NGPSLSDQVLANCNOE---REKLPICTKEGLWTQEPKLCENL 1649  
Db 387 EPLESPVHSGMDSPLRAFOYTNGSFRCAEGFMLRGADIVACDNLGOWTAPAPVCOAL 446  
QY 1650 QGSCPPPSSELNSVEYKCEQGYG---IGAVCS----- 1679  
Db 447 QCDLFPV--NEARVNCSHPFAGAFRYQSVCSFTCNBGLLVGASVLCCLATGNNNSVP 503  
QY 1680 LCVIPSPDPMLPENITADTLEHMMEPVKVQS-----IVCTRRQ 1719  
Db 504 ECAIPCTPLSPONGMTMC---VOPLGSSYSKSTQCFICDEGYSLSGERLDCRSGR 559  
QY 1720 WHPDPVLVHCIOGCEPQAD-GWCDTIINN-----AYCHY 1753  
Db 560 WTDSPWCEAIKCELPAPRQGLDCSDTRGEFNVGSTCFP 600

RESULT 10  
PCT-US91-05059-2  
Sequence 2, Application PC/TUS9105059  
GENERAL INFORMATION:  
APPLICANT: Regents of the Board of the, University of  
APPLICANT: Oklahoma  
TITLE OF INVENTION: Functionally Active Selectin-Derived  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Kilpatrick & Cody  
STREET: 100 Peachtree Street, Suite 3100  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: US  
ZIP: 30303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US91/05059  
FILING DATE: 19910717  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/320408  
FILING DATE: 08-MAR-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/554199  
FILING DATE: 17-JUL-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Pabst, Patrea L.  
REGISTRATION NUMBER: 31,284  
REFERENCE/DOCKET NUMBER: OMREF110CIP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404-572-6508

TELEFAX: 404-572-6555  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 830 amino acids  
TYPE: AMINO ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: YES  
ANTI-SENSE: YES  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE:  
ORGANISM: Homo sapien  
TISSUE TYPE: Blood  
CELL TYPE: Endothelial  
FEATURE:  
NAME/KEY: Disulfide-bond  
LOCATION: 4..25  
FEATURE:  
NAME/KEY: Disulfide-bond  
LOCATION: 60..158  
FEATURE:  
NAME/KEY: Disulfide-bond  
LOCATION: 131..150  
FEATURE:  
NAME/KEY: Disulfide-bond  
LOCATION: 163..174  
FEATURE:  
NAME/KEY: Disulfide-bond  
LOCATION: 168..183  
FEATURE:  
NAME/KEY: Disulfide-bond  
LOCATION: 185..194  
FEATURE:  
NAME/KEY: Disulfide-bond  
LOCATION: 200..244  
FEATURE:  
NAME/KEY: Disulfide-bond  
LOCATION: 213..226  
FEATURE:  
NAME/KEY: Disulfide-bond  
LOCATION: 230..257  
FEATURE:  
NAME/KEY: Disulfide-bond  
LOCATION: 262..306  
FEATURE:  
NAME/KEY: Disulfide-bond  
LOCATION: 275..288  
FEATURE:  
NAME/KEY: Disulfide-bond  
LOCATION: 292..319  
FEATURE:  
NAME/KEY: Disulfide-bond  
LOCATION: 324..368  
FEATURE:  
NAME/KEY: Disulfide-bond  
LOCATION: 337..350  
FEATURE:  
NAME/KEY: Disulfide-bond  
LOCATION: 354..381  
FEATURE:  
NAME/KEY: Disulfide-bond  
LOCATION: 386..430  
FEATURE:  
NAME/KEY: Disulfide-bond  
LOCATION: 399..412  
FEATURE:  
NAME/KEY: Disulfide-bond  
LOCATION: 416..443  
FEATURE:  
NAME/KEY: Disulfide-bond  
LOCATION: 448..492  
FEATURE:  
NAME/KEY: Disulfide-bond  
LOCATION: 461..474  
FEATURE:  
NAME/KEY: Disulfide-bond  
LOCATION: 478..505  
FEATURE:  
NAME/KEY: Disulfide-bond  
LOCATION: 510..554  
FEATURE:  
NAME/KEY: Disulfide-bond  
LOCATION: 523..536  
FEATURE:  
NAME/KEY: Disulfide-bond  
LOCATION: 540..567  
FEATURE:  
NAME/KEY: Disulfide-bond  
LOCATION: 572..616  
FEATURE:  
NAME/KEY: Disulfide-bond  
LOCATION: 585..616  
FEATURE:  
NAME/KEY: Disulfide-bond  
LOCATION: 602..629  
FEATURE:  
NAME/KEY: Disulfide-bond  
LOCATION: 642..686  
FEATURE:  
NAME/KEY: Disulfide-bond  
LOCATION: 655..668  
FEATURE:  
NAME/KEY: Disulfide-bond  
LOCATION: 672..699  
FEATURE:  
NAME/KEY: Disulfide-bond  
LOCATION: 704..748  
FEATURE:  
NAME/KEY: Disulfide-bond  
LOCATION: 717..730  
FEATURE:  
NAME/KEY: Disulfide-bond  
LOCATION: 734..761  
FEATURE:  
NAME/KEY: Binding-site  
LOCATION: 54  
OTHER INFORMATION: /note= "Potential asparagine-linked  
OTHER INFORMATION: glycosylation site"  
FEATURE:  
NAME/KEY: Binding-site  
LOCATION: 98  
OTHER INFORMATION: /note= "Potential asparagine-linked  
OTHER INFORMATION: glycosylation site"  
FEATURE:  
NAME/KEY: Binding-site  
LOCATION: 180  
OTHER INFORMATION: /note= "Potential asparagine-linked  
OTHER INFORMATION: glycosylation site"  
FEATURE:  
NAME/KEY: Binding-site  
LOCATION: 212  
OTHER INFORMATION: /note= "Potential asparagine-linked  
OTHER INFORMATION: glycosylation site"  
FEATURE:  
NAME/KEY: Binding-site  
LOCATION: 219  
OTHER INFORMATION: /note= "Potential asparagine-linked  
OTHER INFORMATION: glycosylation site"  
FEATURE:  
NAME/KEY: Binding-site  
LOCATION: 411  
OTHER INFORMATION: /note= "Potential asparagine-linked  
OTHER INFORMATION: glycosylation site"  
FEATURE:  
NAME/KEY: Binding-site

LOCATION: 460  
OTHER INFORMATION: /note= "Potential asparagine-linked  
OTHER INFORMATION: glycosylation site"  
FEATURE:  
NAME/KEY: Binding-site  
LOCATION: 518  
OTHER INFORMATION: /note= "Potential asparagine-linked  
OTHER INFORMATION: glycosylation site"  
FEATURE:  
NAME/KEY: Binding-site  
LOCATION: 665  
OTHER INFORMATION: /note= "Potential asparagine-linked  
OTHER INFORMATION: glycosylation site"  
FEATURE:  
NAME/KEY: Binding-site  
LOCATION: 716  
OTHER INFORMATION: /note= "Potential asparagine-linked  
OTHER INFORMATION: glycosylation site"  
FEATURE:  
NAME/KEY: Binding-site  
LOCATION: 716  
OTHER INFORMATION: /note= "Potential asparagine-linked  
OTHER INFORMATION: glycosylation site"  
FEATURE:

Query Match 2.5%; Score 243.5; DB 5; Length 830;  
Best Local Similarity 22.0%; Pred. No. 8e-12;  
Matches 111; Conservative 61; Mismatches 163; Indels 169; Gaps 30;

QY 1384 SCIRHPCGKQDSCPSLLLDHADVNTSIGGLMKCAITCQGFALQASSGQYIR-----1438  
DB 162 SCQDMSCSKQGECLRTI-----GNVTC--SCVPGF--YGPEGEYRREGEL 203  
QY 1439 PMQKEILLTS-----SGHW-DQNVSCLPVDCG 1465  
DB 204 ELPOEVLAMNSHPLGNFSPNSQCSFICTDGYQVNGPSKLBELASGIVTNKPPQCLAAQC- 262  
QY 1466 VPPSLVNVANFSC--SEGTFLKRCISIC-----VPPAKQLGSLPWLTCLEDGKMSL 1516  
DB 263 -PPLKIPERGNMCLMSAKAPQWSSCSFSCBEPFALVGEVVO-----CTASGWT 314  
QY 1517 PEVYCK-LAC--DAP-----PIILNANMLLPHCLDNHVGITCKYCKECPGYVA 1563  
DB 315 PAPVCKAVCOHLAPFEGMTDVCVHPLTAF-----YGSCKFEQCPGY--- 358  
QY 1564 ESABGRKRLKILKICLEGIEWO--GSCIPVVEPPPVPEGYECT---NGFSLDSQC 1618  
DB 359 -----RRRG-LDMRCIDSGHWSAPLPCENISCEPESPPWMSDPSLRAFOYDTNC 412  
QY 1619 VLNCNOE---REKLPILCTKESGLMTQEPKLCENIQGCPPEPSLNVEXKCEGVC--- 1672  
DB 413 SFCABEFLMGADIVACDNLGONTAPAPVQALQCDLVP--NEARVNCMPFGAFR 469  
QY 1673 IGAVCSPLCVIPSPDPVWLPEINITADTLE-----HMMN-PVKVQSIIVCTGRQWHPPVL 1726  
DB 470 YGVSCTFTC---NEGILL---VGASVLOCLATGNMNSVPECAIPICT-----PLL 514  
QY 1727 -----VHCIO-----SCEPFQADGWCCTINRAYC-----HYDGDCCSSTLSK 1766  
DB 515 SPONGMTTCVQPLGSSSYKSTCOFICDEGYSLSGPERLICTRSGRMTDPSPMCEATKCP 574  
QY 1767 KVIPIPADCDLDECTGCDPKAEN 1790  
DB 575 LFAPECGSLD-----CSDTRGEFN 593

RESULT 11  
US-08-435-149-3  
Sequence 3, Application US/08435149  
Patent No. 5866402  
GENERAL INFORMATION:  
APPLICANT: INNIS, MICHAEL A.  
APPLICANT: ZAROR, ISABEL  
APPLICANT: CREASEY, ABIA A.  
TITLE OF INVENTION: CHIMERIC MCP AND DAF PROTEINS WITH CELL  
NUMBER OF SEQUENCES: 26  
CORRESPONDENCE ADDRESSES:

ADDRESSEE: CHIRON CORPORATION  
STREET: INTELLECTUAL PROPERTY - R440, P.O. BOX 8097  
CITY: EMERYVILLE  
STATE: CALIFORNIA  
COUNTRY: U.S.A.  
ZIP: 94662-8097  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/435,149  
FILING DATE: 05-MAY-1995  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: SAVEREIDE, PAUL B.  
REGISTRATION NUMBER: 36,914  
REFERENCE/DOCKET NUMBER: 0989,001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (510) 601-2585  
TELEFAX: (510) 655-3542  
TELEX: N/A  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 577 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-435-149-3

Query Match 2.4%; Score 236; DB 2; Length 577;  
Best Local Similarity 23.9%; Pred. No. 2e-11;  
Matches 79; Conservative 38; Mismatches 138; Indels 76; Gaps 13;

QY 1423 CORGFALQASSGQYIRPMQKEILLTSCSSGMDQNV--SCLPV-----DPSLVN 1473  
DB 222 CDGFGYLDGS-----DTVCDSNSTWDPVPCIKVSTDCGLPDPVPAQNALBG 271  
QY 1474 YANFSCSEGTFLKRCISICVPPAKQLGSLPWLTCLEDGKMSLPEVYCKLECDAPPIILN 1533  
DB 272 RTSF--PEDVITYKCESFV--KIPGKDSVYCLKGSQMSDIEFCNRSCEVPTILNS 326  
QY 1534 ANLLPHCLDNHVGITCKYCKECPGYVAESABGRKRLKILKICLEGIEWO--SCI 1591  
DB 327 ASLKQPIITQNYPPVGVVVEYECRPGYRREPSLSP-----KLTCLDNKMSVAVEFC 379  
QY 1592 PVVCEPPPVPEGYECTNFGSLDSQVLNCNQ-----BREKLPILCTKESGLMTQEPK 1645  
DB 380 KKCCPNPGEIRNQIDVPGILFGATISFCNTGYKLPGSTSFCLISGSSVQMSDPLPE 439  
QY 1646 CENLOECPPPEPSELN-----SVEYKEQGYGIGAVCSPLCVI----- 1683  
DB 440 CREIY--CPAPQIDNIIIGERDHYGYSQSVYACNKGMTLGEHSITYCVNNDDEGWS 497  
QY 1684 -PPSD-----PWLPEINITADTLE 1701  
DB 498 GPPECRGKSLTSKVPTVQKPTTVNVPTTE 528

RESULT 12  
US-09-475-460A-32  
Sequence 32, Application US/09475460A  
Patent No. 6316253  
GENERAL INFORMATION:  
APPLICANT: Scott, Elizabeth  
APPLICANT: Innis, Michael  
TITLE OF INVENTION: EXPRESSION VECTORS, TRANSFECTION SYSTEMS, AND METHOD OF  
FILE REFERENCE: 1527,003  
CURRENT APPLICATION NUMBER: US/09/475,460A  
CURRENT FILING DATE: 1999-12-30

	Query March	2.4%	Score 236;	DB 4;	Length 611;
	Best Local Similarity	23.9%;	Pred. No. 2.le-11.		
	Matches	79;	Conservative	38;	Mismatches 136; Indels 76; Gaps 13;
QY	1423 CORGFALQASSGOVIRPWOKEILLTSCSSGHMDONV-SCLPY-----DPSLVN	1473			
Db	256 CDKFPYLIGS-----DTIVODNSSTWDPFPPKLLKXSTDCGLPEPDVPNNQPALGG	305			
QY	1474 YANFSCSEGIYFLKRCSSISCVPAKKQLGSLFWLTLIEDGLMSLPBVCYCKLECDAPPIILN	1533			
Db	306 RTSF--PEDTVITYKCERSFY---KIPEKDSVICLTCKSQSDIEEPFNCRSECVETRLNS	360			
QY	1534 ANLLPHGLDNHNVGTICKYECKRPGYYVAASBEKANVKLLKTLCGLEGGITWGGS--SCI	1591			
Db	361 ASLKOPYITTONYPFGIVAVEYECRGAYREBSLP-----KITCLDNLMKSTAVAECK	413			

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84



**THIS PAGE BLANK (USPTO)**



FT Binding-site 733..743  
 FT /note= "Catalytic zinc binding motif"  
 FT Binding-site 805..809  
 FT /note= "Catalytic zinc binding motif"  
 FT 1396..1459  
 FT /note= "Short consensus repeat (SCR-1)"  
 FT Region 1464..1521  
 FT /note= "Short consensus repeat (SCR-2)"  
 FT Region 1525..1590  
 FT /note= "Short consensus repeat (SCR-3)"  
 FT Region 1595..1646  
 FT /note= "Short consensus repeat (SCR-4)"  
 FT 1653..1729  
 FT /note= "Short consensus repeat (SCR-5)"  
 FT Domain 1733..1758  
 FT /note= "Lin-notch motif (LMR3)"  
 XX MO200232953-A2.  
 XX 25-APR-2002.  
 XX 19-OCT-2001; 2001WO-DK00695.  
 XX 20-OCT-2000; 2000DK-0001571.  
 XX 20-OCT-2000; 2000US-241840P.  
 XX (COMO-) COMO BIOTECH APS.  
 XI Oxvig C, Overgaard MT;  
 XX WPI; 2002-444239/47.  
 DR N-PSDB; AAD38203.  
 XX  
 XX New pregnancy-associated plasma protein A2 and encoding nucleic acids,  
 PT useful as marker for pathological states (e.g. Down's syndrome), or as  
 PT a therapeutic target for drugs that modify the proteolytic activity of  
 PT PAP-A2  
 XX  
 PS Claim 12; Fig 1; 113pp; English.  
 XX  
 CC The invention relates to pregnancy-associated plasma protein A2 (PAP-A2)  
 CC and its corresponding nucleic acid. PAP-A2 is useful as a marker  
 CC for pathological states (e.g. Down's syndrome, acute coronary syndrome,  
 CC unstable angina or myocardial infarction), and is used to treat diseases  
 CC such as restenosis, coronary atherosclerosis, osteoporosis and cancer.  
 CC It is used as a therapeutic target for drugs that modify the proteolytic  
 CC activity of PAP-A2 in pregnant and non-pregnant individuals and it is  
 CC also used in wound healing. The PAP-A2 cDNA is useful for producing  
 CC full length PAP-A2, and the PAP-A2 gene can be used for gene therapy  
 CC to introduce PAP-A2 into target cells. The PAP-A2 proteins can be used  
 CC in the production of antibodies, diagnostic kits, laboratory reagents and  
 CC assays. In the identification of compounds that affect PAP-A2 function;  
 CC as a therapeutic target for the reduction or elimination of IGFBP-5  
 CC proteolytic activity in a cell; and as antigen for immunization. The  
 CC inhibitory agent or enhancing agent is useful in the manufacture of a  
 CC medicament for treating a clinical condition in an individual in need  
 CC of such treatment. The present sequence is human preproPAP-A2 protein.  
 CC PAP-A2 is an active enzyme which cleaves IGFBP-5.  
 XX  
 SQ Sequence 1791 AA;  
 Query Match 100.0%; Score 9856; DB 23; Length 1791;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1791; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 AVEEPAAPWVGDSPIQSGSELLGDDDAVLGNQSRKESLGEAGLQKGSMAATTTTPTTL 180  
 DB 121 AVEEPAAPWVGDSPIQSGSELLGDDDAVLGNQSRKESLGEAGLQKGSMAATTTTPTTL 180  
 QY 181 NEPKETQRGMMAKSRORQVWKRRAEDGGDSGISHPQWPKHSLKHKVKSPPESN 240  
 DB 181 NEPKETQRGMMAKSRORQVWKRRAEDGGDSGISHPQWPKHSLKHKVKSPPESN 240  
 QY 241 QNGEGSYREAFETNSQVGLPIIFYSGRRRLRLRPVLAIEIPREAFTEAMVPEEGON 300  
 DB 241 QNGEGSYREAFETNSQVGLPIIFYSGRRRLRLRPVLAIEIPREAFTEAMVPEEGON 300  
 QY 301 NPATAGVFNDCSHTVSDKMGALGIRSKDGRDARFPFSLCTDRVKATITLISHSRQ 360  
 DB 301 NPATAGVFNDCSHTVSDKMGALGIRSKDGRDARFPFSLCTDRVKATITLISHSRQ 360  
 QY 361 PGTWTHVAATYDGHMALLVYDGTQVASSLDQSGPLNSPFMAASCSLLGGDSSSDGHYFR 420  
 DB 361 PGTWTHVAATYDGHMALLVYDGTQVASSLDQSGPLNSPFMAASCSLLGGDSSSDGHYFR 420  
 QY 421 GHLCGLVFWSTALPQSHFOHSSQHSAGEEATDVLTFASFEVNTVEMVPPREDEKYPRLV 480  
 DB 421 GHLCGLVFWSTALPQSHFOHSSQHSAGEEATDVLTFASFEVNTVEMVPPREDEKYPRLV 480  
 QY 481 LOGFEPEPEILSPLOPPLCGQTVCDNVELISQYNGVPLRGEKVIYQVNNICDDEGLNP 540  
 DB 481 LOGFEPEPEILSPLOPPLCGQTVCDNVELISQYNGVPLRGEKVIYQVNNICDDEGLNP 540  
 QY 541 IVSEEOIRLOHEALNEAFSRYNISWQLSVHQVNSTLRHRVVLVNCPEPKINDHCPEC 600  
 DB 541 IVSEEOIRLOHEALNEAFSRYNISWQLSVHQVNSTLRHRVVLVNCPEPKINDHCPEC 600  
 QY 601 EHPITGYDGGPCRLQGRCSYMNRRDGLCHYECNNMLNDPDDGDCDQVADVKTCEPDP 660  
 DB 601 EHPITGYDGGPCRLQGRCSYMNRRDGLCHYECNNMLNDPDDGDCDQVADVKTCEPDP 660  
 QY 661 SPKRAVMSVKELEKALQLNSTHPLNIYFASVREDLAGAATPWMDKAVTHLGGIVLSPA 720  
 DB 661 SPKRAVMSVKELEKALQLNSTHPLNIYFASVREDLAGAATPWMDKAVTHLGGIVLSPA 720  
 QY 721 YYGMPGHTDMIHVGVGLYHVFVKVSEBESCNDDCKETVPSMEFGDLCADTAPTPKS 780  
 DB 721 YYGMPGHTDMIHVGVGLYHVFVKVSEBESCNDDCKETVPSMEFGDLCADTAPTPKS 780  
 QY 781 ELCEPEPTSPGCFEPFGAPFTNWSYTDNCTDNFTPNQVAKMHCYLDLVYQMTES 840  
 DB 781 ELCEPEPTSPGCFEPFGAPFTNWSYTDNCTDNFTPNQVAKMHCYLDLVYQMTES 840  
 QY 841 RKPPIPIPIPVNIGQTKSLTIHMLPISGVVYDRASGSLGACTEDGTFRQVYHTASR 900  
 DB 841 RKPPIPIPIPVNIGQTKSLTIHMLPISGVVYDRASGSLGACTEDGTFRQVYHTASR 900  
 QY 901 RVCCSSGVTPEBAVGPDDVDQCEPSLOAWSPEVHLVHMNTVPCTEGCSLELLQHP 960  
 DB 901 RVCCSSGVTPEBAVGPDDVDQCEPSLOAWSPEVHLVHMNTVPCTEGCSLELLQHP 960  
 QY 961 VQADTLTLMWTSFPMESSQVLFDTIELLENKESVHLGELDTFCIPILTIKLHVDGKSGV 1020  
 DB 961 VQADTLTLMWTSFPMESSQVLFDTIELLENKESVHLGELDTFCIPILTIKLHVDGKSGV 1020  
 QY 1021 KYVTFDERIEIDALLTSQPSPLCSGCRPVRYOVLNDRPPASGLPVVYTHSHRKFDTVE 1080  
 DB 1021 KYVTFDERIEIDALLTSQPSPLCSGCRPVRYOVLNDRPPASGLPVVYTHSHRKFDTVE 1080  
 QY 1081 VTPGQMTQYVLAAGSELGASPLNHIHGAIPYCGDGKVSERLGEBCDGDGLVSGGCS 1140  
 DB 1081 VTPGQMTQYVLAAGSELGASPLNHIHGAIPYCGDGKVSERLGEBCDGDGLVSGGCS 1140  
 QY 1141 KVCLEBEGFNVGSPSLCYWVEGDCICEPFRKTSIYDCGIYTPKGYLDQWATAVYSHE 1200  
 DB 1141 KVCLEBEGFNVGSPSLCYWVEGDCICEPFRKTSIYDCGIYTPKGYLDQWATAVYSHE 1200  
 QY 1201 DKKCCPVSLVGTGBPHSLICTSYHBDLPNHRPLTGMPCVASENETQDRSEQPEGSLKE 1260



```

Db      1201 DKKKCPVSLTSGEHSILCTSYNHPDLNPHRPLTGMFCVASENETQDRSEQPEGLKKE 1260
Qy      1261 DEWMLKVCNRPGRARAIFPLTTDGLVPGHOOPTVLTLVDGRSNHSLGTGSLSCOH 1320
Db      1261 DEWMLKVCNRPGRARAIFPLTTDGLVPGHOOPTVLTLVDGRSNHSLGTGSLSCOH 1320
Qy      1321 NPLIINTVTHQNVLFHHTTSLVLFSSPRVGISAVALTSTSLGSLAPSNCSISEDEGONH 1380
Db      1321 NPLIINTVTHQNVLFHHTTSLVLFSSPRVGISAVALTSTSLGSLAPSNCSISEDEGONH 1380
Qy      1381 OGOSCHIRPCGKODSCPSLLDBADVNCSTSIGFMKCAITCGRFALQASSGQYIRPM 1440
Db      1441 OKKILLTSSGSHDQNVSCIPVDCGVPDPSLVNANFSCBGTGFKLRCSISCVPAKLQ 1500
Qy      1501 GLSPWLTCLBDGLMSLPEVYCKLECDAPITILNANLLPHCLQDNHVDGTICKYCKRGY 1560
Db      1501 GLSPWLTCLBDGLMSLPEVYCKLECDAPITILNANLLPHCLQDNHVDGTICKYCKRGY 1560
Qy      1561 YVAESAAGKVRNKLKTIQCLEGGIWEQSCIPVCEPPEPVFEGMYECTNGFSLDSQCVL 1620
Db      1561 YVAESAAGKVRNKLKTIQCLEGGIWEQSCIPVCEPPEPVFEGMYECTNGFSLDSQCVL 1620
Qy      1621 NCMQERKPLILCTKESLMQEPKLCENLQCECPPESELSNVYKCEQYIGAVCSPL 1680
Db      1621 NCMQERKPLILCTKESLMQEPKLCENLQCECPPESELSNVYKCEQYIGAVCSPL 1680
Qy      1681 CVIPSPDPVLPENITADTLEHMMEPVKVOSIVCTGRQHPDPVLVHCIOQCEPFOADG 1740
Db      1681 CVIPSPDPVLPENITADTLEHMMEPVKVOSIVCTGRQHPDPVLVHCIOQCEPFOADG 1740
Qy      1741 WCDTINNRAYCHYDGDCCSSTLSSKKVIPPADCDLDECTCRDPAKAEHQ 1791
Db      1741 WCDTINNRAYCHYDGDCCSSTLSSKKVIPPADCDLDECTCRDPAKAEHQ 1791

```

## RESULT 2

```

AAE23826 ID AAE23826 standard; Protein; 1791 AA.
XX
AC AAE23826;
DT 10-SBP-2002 (first entry)
XX
DE Human preproPAP-A2 protein mutant, E734Q.
XX
KM Human; pregnancy-associated plasma protein A2; PAP-A2; Down's syndrome;
KM acute coronary syndrome; angina; myocardial infarction; reestenosis;
KM coronary atherosclerosis; wound healing; osteoporosis; cancer; cadant;
KM gene therapy; IGFBP-5; immunisation; gynaecological; antiinflammatory;
KM vasotropic; nocotropic; cytosstatic; antiarteriosclerotic; vulnerary;
KM osteopathic; enzyme; mutant; mutein.
XX
OS Homo sapiens.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Region 1..233
FT Peptide /note= "Human PAP-A2 prepro region"
FT /label= Signal_peptide
FT /note= "Human PAP-A2 pro region"
FT /note= "Human PAP-A2 protein"
FT /note= "Human mature PAP-A2 protein"
FT /note= "Lin-notch motif (LNR1)"
FT /note= "Lin-notch motif (LNR2)"
FT /note= "Lin-notch motif (LNR3)"
FT Domain

```

```

FT Binding-site 733..743
FT /note= "Catalytic zinc binding motif"
FT Misc-difference 734
FT Binding-site /note= "Wild-type Glu substituted with Gln"
FT /note= "Catalytic zinc binding motif"
FT Region 1396..1459
FT /note= "Short consensus repeat (SCR-1)"
FT Region 1464..1521
FT /note= "Short consensus repeat (SCR-2)"
FT Region 1525..1590
FT /note= "Short consensus repeat (SCR-3)"
FT Region 1595..1646
FT /note= "Short consensus repeat (SCR-4)"
FT Region 1653..1729
FT /note= "Short consensus repeat (SCR-5)"
FT Domain /note= "Lin-notch motif (LNR3)"
FT
XX W0200232953-A2.
XX
XX 25-APR-2002.
XX
XX 19-OCT-2001; 2001WC-DK00695.
XX
XX 20-OCT-2000; 2000DK-0001571.
XX 20-OCT-2000; 2000US-241840P.
XX
XX (COMO-) COMO BIOTECH APS.
XX
XX Oxyig C, Overgaard MT;
XX
XX WPI: 2002-444239/47.
XX N-PSDB; AAD38324.
XX
XX New pregnancy-associated plasma protein A2 and encoding nucleic acids,
XX useful as marker for pathological states (e.g. Down's syndrome), or as
XX a therapeutic target for drugs that modify the proteolytic activity of
XX PAP-A2.
XX
XX Example ; Page -; 113pp; English.
XX
XX The invention relates to pregnancy-associated plasma protein A2 (PAP-A2)
XX and its corresponding nucleic acid. PAP-A2 is useful as a marker
XX for pathological states (e.g. Down's syndrome, acute coronary syndrome,
XX unstable angina or myocardial infarction), and is used to treat diseases
XX such as reestenosis, coronary atherosclerosis, osteoporosis and cancer.
XX It is used as a therapeutic target for drugs that modify the proteolytic
XX activity of PAP-A2 in pregnant and non-pregnant individuals and it is
XX also used in wound healing. The PAP-A2 cDNA is useful for producing
XX full length PAP-A2, and the PAP-A2 gene can be used for gene therapy
XX to introduce PAP-A2 into target cells. The PAP-A2 proteins can be used
XX in the production of antibodies, diagnostic kits, laboratory reagents and
XX assays; in the identification of compounds that affect PAP-A2 function;
XX as a therapeutic target for the reduction or elimination of IGFBP-5
XX proteolytic activity in a cell; and as antigen for immunization. The
XX inhibitory agent or enhancing agent is useful in the manufacture of a
XX medicament for treating a clinical condition in an individual in need
XX of such treatment. The present sequence is human preproPAP-A2 protein
XX mutant. PAP-A2 is an active enzyme which cleaves IGFBP-5.
XX Note: This sequence is not shown in the specification, however it
XX is constructed based on human preproPAP-A2 protein shown in fig 1
XX (AAE23821).
XX
XX Sequence 1791 AA;
XX
Qy Query Match 100.0%; Score 9853; DB 23; Length 1791;
Db Best Local Similarity 99.9%; Pred. No. 0;
Matches 1790; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
1 MMCKIIRISLAIAGALCSANSELGWTTRKSLVEREHUNOVLLLEGRCWLGAKVRRPR 60
1 MMCKIIRISLAIAGALCSANSELGWTTRKSLVEREHUNOVLLLEGRCWLGAKVRRPR 60

```

```

QY 61 ASPOHLEFGVYPSRAGNYLRPYVGEOEIHHHTGRSKPTEGNAVSLVPPDLTENPAGIRG 120
Db 61 ASPOHLEFGVYPSRAGNYLRPYVGEOEIHHHTGRSKPTEGNAVSLVPPDLTENPAGIRG 120
QY 121 AVEBPAPWVGDSPICQSELLGDDDAVIGNQRKSESLGEAGIQKGSAMAAATTTAIFTL 180
Db 121 AVEBPAPWVGDSPICQSELLGDDDAVIGNQRKSESLGEAGIQKGSAMAAATTTAIFTL 180
QY 181 NEPEPEPORGMASRORQVWKRABDQDSCGISSHFOPWPKSLKHKRKSPREESN 240
Db 181 NEPEPEPORGMASRORQVWKRABDQDSCGISSHFOPWPKSLKHKRKSPREESN 240
QY 241 QNGEGSREARETENSQVGLPILYFSGRRELLRPEVLAETIPREAFVEMVPEEGQN 300
Db 241 QNGEGSREARETENSQVGLPILYFSGRRELLRPEVLAETIPREAFVEMVPEEGQN 300
QY 301 NPALIAGVFDCNCHTSDKGMALGIRSGKDKKRDARFFSLCTDRVKKATILISHSRQ 360
Db 301 NPALIAGVFDCNCHTSDKGMALGIRSGKDKKRDARFFSLCTDRVKKATILISHSRQ 360
QY 361 PGTTHVAATYDGRHMLYVPGTOVASLDSGFLNSPFMASCSLLIGDSSSDGHYER 420
Db 361 PGTTHVAATYDGRHMLYVPGTOVASLDSGFLNSPFMASCSLLIGDSSSDGHYER 420
QY 421 GHGLTIVFMSYALPOSHFOHSSSGEERATDLVLTASFEPVNTWVPPRDEKYPRLV 480
Db 421 GHGLTIVFMSYALPOSHFOHSSSGEERATDLVLTASFEPVNTWVPPRDEKYPRLV 480
QY 481 LOGEPEBEILSPIQLCGQTCVNDVLLISQYNGWBLRGEKVIYQVANIODEGLNP 540
Db 481 LOGEPEBEILSPIQLCGQTCVNDVLLISQYNGWBLRGEKVIYQVANIODEGLNP 540
QY 541 IVSEBOQLRQHEALNEASRYNISWQLSVHOVHNSTLHRVYLVCNCESKIGNDHCEC 600
Db 541 IVSEBOQLRQHEALNEASRYNISWQLSVHOVHNSTLHRVYLVCNCESKIGNDHCEC 600
QY 601 EHPPLTGDGDCRLOGRCYSMNRDGLCHVECNMNLNFDGDDCCDPVAVRKTCTFPD 660
Db 601 EHPPLTGDGDCRLOGRCYSMNRDGLCHVECNMNLNFDGDDCCDPVAVRKTCTFPD 660
QY 661 SPKRAYMSVKELKALQLNSTHPLNITYPASSVREDLAAGATWPKDAVTHLGGIVLSPA 720
Db 661 SPKRAYMSVKELKALQLNSTHPLNITYPASSVREDLAAGATWPKDAVTHLGGIVLSPA 720
QY 721 YYGMPGHTDTHIHGVHGLGYHVFKYGSERESCHDPCKEIVPSMETGDLCAADTAPTPKS 780
Db 721 YYGMPGHTDTHIHGVHGLGYHVFKYGSERESCHDPCKEIVPSMETGDLCAADTAPTPKS 780
QY 781 ELCEPEPTSDTCGFTFRPAGPFTNYMSYTDNCTDNFTPNQVAMHCYLDLVYQOMTES 840
Db 781 ELCEPEPTSDTCGFTFRPAGPFTNYMSYTDNCTDNFTPNQVAMHCYLDLVYQOMTES 840
QY 841 RKPTPIPIPMVNIQGTNKSLLTIHMLPISGVVYDRASGLCGACTEDGTFRQYVTAASR 900
Db 841 RKPTPIPIPMVNIQGTNKSLLTIHMLPISGVVYDRASGLCGACTEDGTFRQYVTAASR 900
QY 901 RVCSSSGVTFBEAVGPPDVQPCPSLQAMSPEVHLIYHMMYVPCPTGCSLEBLFQHP 960
Db 901 RVCSSSGVTFBEAVGPPDVQPCPSLQAMSPEVHLIYHMMYVPCPTGCSLEBLFQHP 960
QY 961 VOADTLTMTSFPMESSQVLFDTIELLENKESVHLGPLDFCDPLITIKLHVQKASGV 1020
Db 961 VOADTLTMTSFPMESSQVLFDTIELLENKESVHLGPLDFCDPLITIKLHVQKASGV 1020
QY 1021 KYVTFDERIEIDALILTSQPHSPLCSGCRPVRYQLRDPFPFASGLPVVYVTHSHRFTDVE 1080
Db 1021 KYVTFDERIEIDALILTSQPHSPLCSGCRPVRYQLRDPFPFASGLPVVYVTHSHRFTDVE 1080
QY 1081 VTPEGMYQVYLAENGELGEASPLNIHIGAPYGGQKVSERLCEBCDDGLVSGDCS 1140
Db 1081 VTPEGMYQVYLAENGELGEASPLNIHIGAPYGGQKVSERLCEBCDDGLVSGDCS 1140

```

```

QY 1141 KVCELEBGFNCVGPSPSLCYMVEGDGICEPERKTSIVDCGIYTPKGYLDQMATAVSSHE 1200
Db 1141 KVCELEBGFNCVGPSPSLCYMVEGDGICEPERKTSIVDCGIYTPKGYLDQMATAVSSHE 1200
QY 1201 DKKKCPVSLVTGEPHSLICTSYHNPDLFNHRPLTGMFPCVASENETODRSQBPGLSKE 1260
Db 1201 DKKKCPVSLVTGEPHSLICTSYHNPDLFNHRPLTGMFPCVASENETODRSQBPGLSKE 1260
QY 1261 DEWMLKVCNRPGRARAFIPLTTDGLVPGHHOQPTVTLVTVDRGSNHSIGTGLSCQH 1320
Db 1261 DEWMLKVCNRPGRARAFIPLTTDGLVPGHHOQPTVTLVTVDRGSNHSIGTGLSCQH 1320
QY 1321 NPLIINTVHQNVLFFHTTSLVNLNFSPPRGVISAVALRTSRIGLSAPSNCISDEBQNH 1380
Db 1321 NPLIINTVHQNVLFFHTTSLVNLNFSPPRGVISAVALRTSRIGLSAPSNCISDEBQNH 1380
QY 1381 QGOSCHIRPCGKODSCPSLLBDHADVNVCTSIGFGLMKCAITCORGFALQASSQYIRPM 1440
Db 1381 QGOSCHIRPCGKODSCPSLLBDHADVNVCTSIGFGLMKCAITCORGFALQASSQYIRPM 1440
QY 1441 QKEILLTCSGSHMDQNVSCLPVDCGVPDPSLVNYANFSCSBEKTEFLRCSISCVPAKLQ 1500
Db 1441 QKEILLTCSGSHMDQNVSCLPVDCGVPDPSLVNYANFSCSBEKTEFLRCSISCVPAKLQ 1500
QY 1501 GLSPMLTCLBEDGLMSLEBYVCKLECDAPRIILNANLLPHCLQDNHVGITCKYECKRGY 1560
Db 1501 GLSPMLTCLBEDGLMSLEBYVCKLECDAPRIILNANLLPHCLQDNHVGITCKYECKRGY 1560
QY 1561 YVASAEGKYNKLLKIQCLEGGIWEQSCIPVCEPPPVFEGMYECTNGFSLDSQCVL 1620
Db 1561 YVASAEGKYNKLLKIQCLEGGIWEQSCIPVCEPPPVFEGMYECTNGFSLDSQCVL 1620
QY 1621 NCNOREKPLTCTKEGLMTOEPFLCENLQECPPPSSELSVYKCEOGYIGAVCSPL 1680
Db 1621 NCNOREKPLTCTKEGLMTOEPFLCENLQECPPPSSELSVYKCEOGYIGAVCSPL 1680
QY 1681 CVIPSPDPMLENITADTLEHMMEPYKVOISVCTGRHMHDPVLVHCIOSECPFOADG 1740
Db 1681 CVIPSPDPMLENITADTLEHMMEPYKVOISVCTGRHMHDPVLVHCIOSECPFOADG 1740
QY 1741 WCDTINNRAYCHYDGDCCSSTLSSKVIYPPADCDLDECTCRDPKABENQ 1791
Db 1741 WCDTINNRAYCHYDGDCCSSTLSSKVIYPPADCDLDECTCRDPKABENQ 1791

RESULT 3
ABB84488
ID ABB84488 standard; Protein; 1791 AA.
AC ABB84488;
DT 24-DEC-2002 (first entry)
XX
XX Human pregnancy associated protein PAP-Ea SEQ ID 3.
DE
XX PAP-E; human; pregnancy associated plasma protein E; abortive;
XX contraceptive; gene therapy; vaccine; pregnancy; antenatal; diagnosis;
XX dyegenetic pregnancy.
OS Homo sapiens.
XX
XX US2002102252-A1.
XX
XX 01-AUG-2002.
XX
XX 06-APR-2001; 2001US-0827998.
XX
XX 26-MAY-2000; 2000US-207456P.
XX
XX (GUTY/) GU Y.
XX PA (SHAN/) SHANNON M E.
XX
XX Gu Y, Shannon ME;
PI

```

XX WPI; 2002-697817/75.  
 DR N-PSDB; ABS74483, ABS74484.  
 XX  
 PT New isolated nucleic acid encoding an isoform of human pregnancy  
 associated plasma protein E, for preventing or aborting pregnancy  
 XX  
 PS Claim 1; Figure 3; 353bp; English.  
 XX  
 CC This invention describes a novel isolated nucleic acid that encodes  
 CC one of three new isoforms of human pregnancy associated plasma protein E,  
 CC hPAP-E. The products of the invention have abortive and contraceptive  
 CC activity. and can be used for gene therapy or in a vaccine. The nucleic  
 CC acid, polypeptide encoded by it, or antibody to the polypeptide can be  
 CC used in pharmaceutical compositions or vaccines for preventing or  
 CC aborting pregnancy. PAP-E is used in the antenatal diagnosis of  
 CC dysgenetic pregnancies. The nucleic acids are used as probes to assess  
 CC the level of PAP-E isoform mRNA in chorionic villus samples, and the  
 CC antibodies can be used to assess the expression levels of PAP-E isoform  
 CC proteins in chorionic villus samples, to diagnose dysgenetic pregnancies  
 CC antenatally. This sequence represents a human PAP-E protein described in  
 CC the disclosure of the invention.

XX Sequence 1791 AA;

Query Match. 99.8%; Score 9836; DB 23; Length 1791;

Best Local Similarity 99.8%; Pred. No. 0;  
 Matches 1788; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MMCKILIRISLAIAGALCSANSELGMRKSLVEREHLNOVLLEGERCMLGAKVRPR 60  
 DB 1 MMCKILIRISLAIAGALCSANSELGMRKSLVEREHLNOVLLEGERCMLGAKVRPR 60  
 QY 61 ASPOHHLFGYVPSAGNYLRLPYGEOIHHTGSKPDTEGNANSLVPPDLTENPAGLRG 120  
 DB 61 ASPOHHLFGYVPSAGNYLRLPYGEOIHHTGSKPDTEGNANSLVPPDLTENPAGLRG 120  
 QY 121 AVEEAPAPWVDSPFISQSELLGDDAYLGNORSKESLGEAGIIOKGSAMATTTTAITTL 180  
 DB 121 AVEEAPAPWVDSPFISQSELLGDDAYLGNORSKESLGEAGIIOKGSAMATTTTAITTL 180  
 QY 121 AVEEAPAPWVDSPFISQSELLGDDAYLGNORSKESLGEAGIIOKGSAMATTTTAITTL 180  
 DB 121 AVEEAPAPWVDSPFISQSELLGDDAYLGNORSKESLGEAGIIOKGSAMATTTTAITTL 180  
 QY 181 NEPKETORRGMAKSRQRRQYKRAEDGQDSGISHPQWPMGLSKHRYKSPRESN 240  
 DB 181 NEPKETORRGMAKSRQRRQYKRAEDGQDSGISHPQWPMGLSKHRYKSPRESN 240  
 QY 241 QNGEGGYRAETNSQVGPILYFSGRRELLRPEVLAIPREATTVEAMVKEBEGON 300  
 DB 241 QNGEGGYRAETNSQVGPILYFSGRRELLRPEVLAIPREATTVEAMVKEBEGON 300  
 QY 241 QNGEGGYRAETNSQVGPILYFSGRRELLRPEVLAIPREATTVEAMVKEBEGON 300  
 DB 241 QNGEGGYRAETNSQVGPILYFSGRRELLRPEVLAIPREATTVEAMVKEBEGON 300  
 QY 301 NPAILIAGVFNCSHTVSDKGMALGIRSGKDKRDAFPFSLCTDRYKKAATILISHRYQ 360  
 DB 301 NPAILIAGVFNCSHTVSDKGMALGIRSGKDKRDAFPFSLCTDRYKKAATILISHRYQ 360  
 QY 301 NPAILIAGVFNCSHTVSDKGMALGIRSGKDKRDAFPFSLCTDRYKKAATILISHRYQ 360  
 DB 301 NPAILIAGVFNCSHTVSDKGMALGIRSGKDKRDAFPFSLCTDRYKKAATILISHRYQ 360  
 QY 361 PGTWTHAATYDGRHMLLVYDGTQVASSLDQSGPLNSPFMAKSCSLILGGDSSEGHYFR 420  
 DB 361 PGTWTHAATYDGRHMLLVYDGTQVASSLDQSGPLNSPFMAKSCSLILGGDSSEGHYFR 420  
 QY 421 GHGLTVLFWNSTALPQSHRQSSSHSSEBEATDLVLTASFEPTVTEWVPFDEKRYPLEV 480  
 DB 421 GHGLTVLFWNSTALPQSHRQSSSHSSEBEATDLVLTASFEPTVTEWVPFDEKRYPLEV 480  
 QY 481 LOGEPEPELISPLQPLCGQTVCDNYELISQYNGYMPRLGEXYIRYOVNI CDDEGLNP 540  
 DB 481 LOGEPEPELISPLQPLCGQTVCDNYELISQYNGYMPRLGEXYIRYOVNI CDDEGLNP 540  
 QY 541 IYSEEOIRLOHEALNEAFSRYNISWOLSVHGVNSTLRHRVVLVNCBPSKIGNDHCPBC 600  
 DB 541 IYSEEOIRLOHEALNEAFSRYNISWOLSVHGVNSTLRHRVVLVNCBPSKIGNDHCPBC 600  
 QY 601 EHLPLTGDDGDCRLQGRCTYWRNRDGLCHVECNMLNDFDGDCCDPQVADVRKTCFDPD 660  
 DB 601 EHLPLTGDDGDCRLQGRCTYWRNRDGLCHVECNMLNDFDGDCCDPQVADVRKTCFDPD 660

QY 661 SPKRAYMSVKELKALQLNSTHFLNIYPASSVREDIAGAATWMDKDAVTHLGIVLSPA 720  
 DB 661 SPKRAYMSVKELKALQLNSTHFLNIYPASSVREDIAGAATWMDKDAVTHLGIVLSPA 720  
 QY 721 YGMPGHTDPMHEGVHVLGLYHFKVSEBESCNBPCKETVSMETGDLCAOTAPPKS 780  
 DB 721 YGMPGHTDPMHEGVHVLGLYHFKVSEBESCNBPCKETVSMETGDLCAOTAPPKS 780  
 QY 781 ELCREPEPSDTCGTFRPGAPFTNYSYTDNCTNFTFNQVARMCYLDLYOQMTES 840  
 DB 781 ELCREPEPSDTCGTFRPGAPFTNYSYTDNCTNFTFNQVARMCYLDLYOQMTES 840  
 QY 841 RKPTPIPIPMVIGQTNKSLTIIMLPISGVVYDRASGSLCAGCTEDGTRQYVHTASR 900  
 DB 841 RKPTPIPIPMVIGQTNKSLTIIMLPISGVVYDRASGSLCAGCTEDGTRQYVHTASR 900  
 QY 901 RVCDSGYWTPBEAVGPPVDQPCBPSLQMSPEVHLYHNMWTVPCTEGCSLELLFQHP 960  
 DB 901 RVCDSGYWTPBEAVGPPVDQPCBPSLQMSPEVHLYHNMWTVPCTEGCSLELLFQHP 960  
 QY 961 VOADTLTMTSPFMESSQVLPTEILLKESVHLCGLDTPCDIPLTLIHYDGVKSGV 1020  
 DB 961 VOADTLTMTSPFMESSQVLPTEILLKESVHLCGLDTPCDIPLTLIHYDGVKSGV 1020  
 QY 1021 KYVTFDERIEIDALLTSQPHSLCSGCRPVRYQVLRDPPASGLPVVYTHSHRKFTDVE 1080  
 DB 1021 KYVTFDERIEIDALLTSQPHSLCSGCRPVRYQVLRDPPASGLPVVYTHSHRKFTDVE 1080  
 QY 1081 VTPGQMYOVVLAEGAGLEASPLNHIHGAPYCGDKVYSERLGEBCDDGLVSGDGS 1140  
 DB 1081 VTPGQMYOVVLAEGAGLEASPLNHIHGAPYCGDKVYSERLGEBCDDGLVSGDGS 1140  
 QY 1141 KVOLEBEGFNCVGEPSLCVYWEEDGICEPERRTSYVDCGIIYTPKGLDMATRAYSHE 1200  
 DB 1141 KVOLEBEGFNCVGEPSLCVYWEEDGICEPERRTSYVDCGIIYTPKGLDMATRAYSHE 1200  
 QY 1201 DKKKCPVSLVTGEPHSLICTSYHAPDLPHNRPLTGMFPCVASENETODRSEOGESLKE 1260  
 DB 1201 DKKKCPVSLVTGEPHSLICTSYHAPDLPHNRPLTGMFPCVASENETODRSEOGESLKE 1260  
 QY 1261 DEWLVKVCENRPGEBARAFILFTTDLVPGEHQOPTVLTLYTVRGSNSHSLGYGLSCQH 1320  
 DB 1261 DEWLVKVCENRPGEBARAFILFTTDLVPGEHQOPTVLTLYTVRGSNSHSLGYGLSCQH 1320  
 QY 1321 NPILINVTHTQNLVFNHTTSVLANFSRPGISAVALRTSSRIGLSPNSCISEDEQNH 1380  
 DB 1321 NPILINVTHTQNLVFNHTTSVLANFSRPGISAVALRTSSRIGLSPNSCISEDEQNH 1380  
 QY 1381 QGOSCIHRPCGKODSCPSLLDHDADVNCISIGBLMKCAITCORGFALQASSGQYIRPM 1440  
 DB 1381 QGOSCIHRPCGKODSCPSLLDHDADVNCISIGBLMKCAITCORGFALQASSGQYIRPM 1440  
 QY 1441 QKSIILLTSSGSHWDQVNSCLPVDGVPDPSLVNVAANSCEGRTFLKRCGISICVPPAKLO 1500  
 DB 1441 QKSIILLTSSGSHWDQVNSCLPVDGVPDPSLVNVAANSCEGRTFLKRCGISICVPPAKLO 1500  
 QY 1501 GLSPMLTCLLEDGLMSLPYVYCKLECDAPRIILNANLLPHCLDNDHVGITCKYECRGY 1560  
 DB 1501 GLSPMLTCLLEDGLMSLPYVYCKLECDAPRIILNANLLPHCLDNDHVGITCKYECRGY 1560  
 QY 1561 YVABASAGVRYNKLKIQCLEGGIWEQSGCIPVCEPFPVFEAGMECTNGFSLDSQCVL 1620  
 DB 1561 YVABASAGVRYNKLKIQCLEGGIWEQSGCIPVCEPFPVFEAGMECTNGFSLDSQCVL 1620  
 QY 1621 NCGOERREKLPIILCTKGLMTQBEKLCENTQSGCPRPPELSNVEYKCEOGYIGAYVSP 1680  
 DB 1621 NCGOERREKLPIILCTKGLMTQBEKLCENTQSGCPRPPELSNVEYKCEOGYIGAYVSP 1680  
 QY 1681 CVIPSPDPVWLBNITADTLEHMMEPYKQSYICTGRQNHPPVYLHNCIQSGEPFADG 1740  
 DB 1681 CVIPSPDPVWLBNITADTLEHMMEPYKQSYICTGRQNHPPVYLHNCIQSGEPFADG 1740  
 QY 1741 WCDTINNRAYCHVDGDCSSSTLSKKVIFPADCDLDECTCRDPKAEBO 1791

Db 1741 WCDTNNRAYCHYDGDCCSSTLSSKKVTFPAADCDLDECTCRDPKAEHQ 1791

RESULT 4  
ABB84490  
ID ABB84490 standard; Protein; 1770 AA.

XX ABB84490;

XX 24-DEC-2002 (first entry)

XX Human pregnancy associated protein PAP-Eb SEQ ID 10.

XX PAP-E; human; pregnancy associated plasma protein E; abortive;

XX contraceptive; gene therapy; vaccine; pregnancy; antenatal; diagnosis;

XX dyogenetic pregnancy.

XX Homo sapiens.

XX US2002102252-A1.

XX 01-AUG-2002.

XX 06-APR-2001; 2001US-0827998.

XX 26-MAY-2000; 2000US-207456P.

XX (GUYV/) GU Y.

XX (SHAN) SHANNON M E.

XX Gu Y, Shannon ME;

XX WPI; 2002-697817/75.

XX N-PSDB; ABS74488; ABS74489.

XX New isolated nucleic acid encoding an isoform of human pregnancy

XX associated plasma protein E, for preventing or aborting pregnancy -

XX Claim 8; Figure 4; 353pp; English.

XX This invention describes a novel isolated nucleic acid that encodes

XX one of three new isoforms of human pregnancy associated plasma protein E,

XX hPAP-E. The products of the invention have abortive and contraceptive

XX activity and can be used for gene therapy or in a vaccine. The nucleic

XX acid, polypeptide encoded by it, or antibody to the polypeptide can be

XX used in pharmaceutical compositions or vaccines for preventing or

XX aborting pregnancy. PAP-E is used in the antenatal diagnosis of

XX dyogenetic pregnancies. The nucleic acids are used as probes to assess

XX the level of PAP-E isoform mRNA in chorionic villus samples, and the

XX antibodies can be used to assess the expression levels of PAP-E isoform

XX proteins in chorionic villus samples, to diagnose dyogenetic pregnancies

XX antenatally. This sequence represents a human PAP-E protein described in

XX the disclosure of the invention.

SO Sequence 1770 AA;

Query Match 96.5%; Score 9507; DB 23; Length 1770;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 1732; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MMCLKILIRISLAIAGMALGANSSELGTRKKSILVERHILNOVLEGRGWLGAQVRRP 60

Db 1 MMCLKILIRISLAIAGMALGANSSELGTRKKSILVERHILNOVLEGRGWLGAQVRRP 60

QY 61 ASPQHLHFGVPSRAGNYLRPYVGEDEIHHTGSKSPTEGNAVSLVPPDLTENPAGLRG 120

Db 61 ASPQHLHFGVPSRAGNYLRPYVGEDEIHHTGSKSPTEGNAVSLVPPDLTENPAGLRG 120

QY 121 AVEBPAPWVDSPIGOSELLGDDDAYIGNORSKESLGEAGI OKGSAMAATTTAIFPTTL 180

Db 121 AVEBPAPWVDSPIGOSELLGDDDAYIGNORSKESLGEAGI OKGSAMAATTTAIFPTTL 180

QY 181 NEPKETQRGMAKSRQRQVWKRAEDGOGDSGSISSHFOQWPXGSLKXHVKSPPRESN 240

Db 181 NEPKETQRGMAKSRQRQVWKRAEDGOGDSGSISSHFOQWPXGSLKXHVKSPPRESN 240

QY 241 QNGGSSYREAFETFNQVGLPILYFSGRREBLIRPEVLAIPREAFTEAVKPEEGON 300

Db 241 QNGGSSYREAFETFNQVGLPILYFSGRREBLIRPEVLAIPREAFTEAVKPEEGON 300

QY 301 NPATIAVFNCSHTVBDKMGALGIRSGKDKGRDARFPFSLCTDRKXATILISHRYQ 360

Db 301 NPATIAVFNCSHTVBDKMGALGIRSGKDKGRDARFPFSLCTDRKXATILISHRYQ 360

QY 361 PGTWTHAAATYDGRHMLLYVDGTQVASSLDQSGPLNSPFMAACSLVLGGDSSEGHYFR 420

Db 361 PGTWTHAAATYDGRHMLLYVDGTQVASSLDQSGPLNSPFMAACSLVLGGDSSEGHYFR 420

QY 421 GHGLTVFWSTALFQSHFQHSQHSSEBEATDLVLTASFEVNTVAVPPREDEKYPRLV 480

Db 421 GHGLTVFWSTALFQSHFQHSQHSSEBEATDLVLTASFEVNTVAVPPREDEKYPRLV 480

QY 481 LQGFPEPELILSPLOPLCGQTVCDNVELISQVNGVPLRGEKIRYQVNICDDEGLNP 540

Db 481 LQGFPEPELILSPLOPLCGQTVCDNVELISQVNGVPLRGEKIRYQVNICDDEGLNP 540

QY 541 IYSEEQIRLOHEALNEAFSRYNISWQSLVQVHNSTLRHVVLVNCEPSKIGNDHCDPEC 600

Db 541 IYSEEQIRLOHEALNEAFSRYNISWQSLVQVHNSTLRHVVLVNCEPSKIGNDHCDPEC 600

QY 601 EHPITGYDGDRCRLOGRCYSMNRDGLCHVECNMMLNDPDDGDCDPOVADVRKTCFDP 660

Db 601 EHPITGYDGDRCRLOGRCYSMNRDGLCHVECNMMLNDPDDGDCDPOVADVRKTCFDP 660

QY 661 SPRKAVMSVEIKKALDNLNSTHPLNITPASFVREDLAGAATWPDKAVVHLGSLVSPA 720

Db 661 SPRKAVMSVEIKKALDNLNSTHPLNITPASFVREDLAGAATWPDKAVVHLGSLVSPA 720

QY 721 YYGMPGHTDIMEHVGVLGLYHFKGVSRERSCNDPCKETVSMETGDLCAADTAPTPKS 780

Db 721 YYGMPGHTDIMEHVGVLGLYHFKGVSRERSCNDPCKETVSMETGDLCAADTAPTPKS 780

QY 781 ELCEPEPTSDTCGCTFPFGAPFTNWSYTDNCTDNFTNOVARNACIYDLVYQMTES 840

Db 781 ELCEPEPTSDTCGCTFPFGAPFTNWSYTDNCTDNFTNOVARNACIYDLVYQMTES 840

QY 841 RKPIPIIPVVGQTKSLTIHMLPISGVVDVDRASGLCGACTEDGTROQVHTASSR 900

Db 841 RKPIPIIPVVGQTKSLTIHMLPISGVVDVDRASGLCGACTEDGTROQVHTASSR 900

QY 901 RVCDSSGYWTPBEAVGPDDVQPCPSLOAWSPEVHLYHNMNTVPCPTBGCSLELFRQP 960

Db 901 RVCDSSGYWTPBEAVGPDDVQPCPSLOAWSPEVHLYHNMNTVPCPTBGCSLELFRQP 960

QY 961 VQADTLTLWVTSFMESSOVLFTTEILLENKESVHLGRLDTFCDIPLTIKLVNDGKXSV 1020

Db 961 VQADTLTLWVTSFMESSOVLFTTEILLENKESVHLGRLDTFCDIPLTIKLVNDGKXSV 1020

QY 1021 KVTTFDEIRIETDALLTSOPHSPLCSGCRPVRYOVLADPPASGLPVVYVHSHKFTDVE 1080

Db 1021 KVTTFDEIRIETDALLTSOPHSPLCSGCRPVRYOVLADPPASGLPVVYVHSHKFTDVE 1080

QY 1081 VTPGQMTQYVLAAGSELGEASPLNHIHGAAPCGDGKYSERLGECECDGDLVSGGCS 1140

Db 1081 VTPGQMTQYVLAAGSELGEASPLNHIHGAAPCGDGKYSERLGECECDGDLVSGGCS 1140

QY 1141 KVCLEBEGFNCVGBPSLCYMEBGDICEPPERKTSIVDCGIYTPKGYLDQWATRAYSHE 1200

Db 1141 KVCLEBEGFNCVGBPSLCYMEBGDICEPPERKTSIVDCGIYTPKGYLDQWATRAYSHE 1200

QY 1201 DKKCCPVSLVTGEPHSLICTSYHNDLPNHRPLTWCFVASENTOODRSEOPGSLIKE 1260

Db 1201 DKKCCPVSLVTGEPHSLICTSYHNDLPNHRPLTWCFVASENTOODRSEOPGSLIKE 1260

QY 1261 DEVMKVCENRPGBARAIFLFTLDGLVPSBHQQPTVTLVLTVDVRSGNHSLGTGSLCOH 1320

Db 1261 DEVLKVCFRPGARAFITFLITDGLVGBHQPIVTLTLTVDRGSHSLGTGLSCQH 1320  
 Qy 1321 NPILINVTHTONVLFHHTTVLNFSSPRVIGISVALRTSRIGLSPSNCISEDEQNH 1380  
 Db 1321 NPILINVTHTONVLFHHTTVLNFSSPRVIGISVALRTSRIGLSPSNCISEDEQNH 1380  
 Qy 1381 QGQSCIRHPCGKODSCPSLLLDHADVNCTSIGGLMKCAITCORGALQASGQYIRPM 1440  
 Db 1381 QGQSCIRHPCGKODSCPSLLLDHADVNCTSIGGLMKCAITCORGALQASGQYIRPM 1440  
 Qy 1441 QKXIIILTCSSGHDQNVSCIPVDCGVPDPSLVNANFSCSEGTFLKRCISICVPAKLQ 1500  
 Db 1441 QKXIIILTCSSGHDQNVSCIPVDCGVPDPSLVNANFSCSEGTFLKRCISICVPAKLQ 1500  
 Qy 1501 GLSPMLTCLDEGJMSLPEVYCKLECDAPITILNANLLPHCLDNHVGITCKYECKRGY 1560  
 Db 1501 GLSPMLTCLDEGJMSLPEVYCKLECDAPITILNANLLPHCLDNHVGITCKYECKRGY 1560  
 Qy 1561 YVAESAAGKYRNKLLKIQCLEGGIWEQSGCIPVCEPPEPPPEGMVETNGFSLDSQCVL 1620  
 Db 1561 YVAESAAGKYRNKLLKIQCLEGGIWEQSGCIPVCEPPEPPPEGMVETNGFSLDSQCVL 1620  
 Qy 1621 NCNQRERKLPILCTKEGLMTQEFKLCENLQGECPPEPSELINSVEYKCEGQYIGAVCSPL 1680  
 Db 1621 NCNQRERKLPILCTKEGLMTQEFKLCENLQGECPPEPSELINSVEYKCEGQYIGAVCSPL 1680  
 Qy 1681 CVLPSPSPVNLPEKITADTLEHMMERPKVQSIYCTGRRQWHPDVLVHICQSC 1734  
 Db 1681 CVLPSPSPVNLPEKITADTLEHMMERPKVQSIYCTGRRQWHPDVLVHICQSC 1734

## RESULT 5

ID ABB04485 standard; Protein; 1624 AA.  
 AC ABB04485;  
 DT 14-MAY-2002 (first entry)  
 DE Human pregnancy-associated plasma protein E.  
 KM Human; pregnancy-associated plasma protein E; PAPP-B; trisomy 13;  
 OS Homo sapiens.  
 PN WO200195855-A2.  
 PD 20-DEC-2001.  
 PF 15-JUN-2001; 2001WO-EP06831.  
 PR 16-JUN-2000; 2000DE-1028971.  
 PA (TSCHECH) TSCHECH H.  
 PA (FARR) FARR M.  
 PI Tschesche H, Farr M;  
 DR WPI; 2002-098007/13.  
 DR N-PSDB; ABA05845.  
 PT New DNA encoding pregnancy-associated plasma protein-E, useful for  
 PT diagnosis of trisomy, also related proteins and antibodies -  
 PS Claim 6; Page 22; 31pp; German.  
 CC The present invention provides the protein and coding sequences of human  
 CC pregnancy-associated plasma protein-E (PAPP-E). The sequences can be used  
 CC in human genetic diagnosis, especially for detecting trisomy 13, 18, 21  
 CC or 22. The present sequence is the protein of the invention.

SQ Sequence 1624 AA;  
 Query Match 90.8%; Score 8945; DB 23; Length 1624;  
 Best Local Similarity 99.8%; Pred. No. 0;  
 Matches 1620; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
 Qy 168 MAATTTAFTTLNEKPEFORGRMAKSRORQVWRKRRADGCGDGSISHPQWPKHSL 227  
 Db 1 MAATTTAFTTLNEKPEFORGRMAKSRORQVWRKRRADGCGDGSISHPQWPKHSL 60  
 Qy 228 KHRVYKSPPEESNQNGEGSYREAEFTNSQVGLPIIYFSGRRLRLRPEVLAEIPREAF 287  
 Db 61 KHRVYKSPPEESNQNGEGSYREAEFTNSQVGLPIIYFSGRRLRLRPEVLAEIPREAF 140  
 Qy 288 TYEAWYKPEEGQNNPAILIAGVFNCSHTVSDKMAIGISGKRGKRDARFFSLCTDRV 347  
 Db 121 TYEAWYKPEEGQNNPAILIAGVFNCSHTVSDKMAIGISGKRGKRDARFFSLCTDRV 180  
 Qy 348 KKATILISHSRVOPGTMTHVAATYDGRHMLYVDGTQVASSLDSQSGPLNSPFMA5CRSL 407  
 Db 181 KKATILISHSRVOPGTMTHVAATYDGRHMLYVDGTQVASSLDSQSGPLNSPFMA5CRSL 240  
 Qy 408 LGDSSDEGHYFRGLGTLVFNSTALPQSHFOHSSQHS5GEEBATDLVLTASFEPVNTW 467  
 Db 241 LGDSSDEGHYFRGLGTLVFNSTALPQSHFOHSSQHS5GEEBATDLVLTASFEPVNTW 300  
 Qy 468 VPRDEKYPRLVLOGEPEPELISLOPLCGQVYCDNVELISQYNGVWPLAGEKYIRY 527  
 Db 301 VPRDEKYPRLVLOGEPEPELISLOPLCGQVYCDNVELISQYNGVWPLAGEKYIRY 360  
 Qy 528 QVANI CDDEGLNPIVSEBQIRLOHEALNEAFSYNLSWQSVYQVNSTLRHVLVNC 587  
 Db 361 QVANI CDDEGLNPIVSEBQIRLOHEALNEAFSYNLSWQSVYQVNSTLRHVLVNC 420  
 Qy 588 PSKIGNDHCDPECEHPLTGYDGDCLQGRCY5WNRDGLCHVECNMNLDPDGDCCDP 647  
 Db 421 PSKIGNDHCDPECEHPLTGYDGDCLQGRCY5WNRDGLCHVECNMNLDPDGDCCDP 480  
 Qy 648 QVADVAKTCGDPSPKRAYSVKELKEALQNSTHRLNIFPASYVEDLAGAATWPMKD 707  
 Db 481 QVADVAKTCGDPSPKRAYSVKELKEALQNSTHRLNIFPASYVEDLAGAATWPMKD 540  
 Qy 708 AVTHLGIIVLSPAYGMPGHTDMHVEGVGLLYVFKGVSRESCNDPCKETVPSMET 767  
 Db 541 AVTHLGIIVLSPAYGMPGHTDMHVEGVGLLYVFKGVSRESCNDPCKETVPSMET 600  
 Qy 768 GDLCAADTAPPKSELCREBEPTSDTCGTRFPGAPFTNWSYTDNCTDNFTNOVARMH 827  
 Db 601 GDLCAADTAPPKSELCREBEPTSDTCGTRFPGAPFTNWSYTDNCTDNFTNOVARMH 660  
 Qy 828 CYLDLYYQWTEBKRKTPPIIPMVYQGNKSLTIMLPIISGVYVDRASGSLCGACTED 887  
 Db 661 CYLDLYYQWTEBKRKTPPIIPMVYQGNKSLTIMLPIISGVYVDRASGSLCGACTED 720  
 Qy 888 GTFROYVHTASSRRVCDSSGWTPEEAVGPVDVDCPEB5LOAMSP5EVLHYHNMVPCP 947  
 Db 721 GTFROYVHTASSRRVCDSSGWTPEEAVGPVDVDCPEB5LOAMSP5EVLHYHNMVPCP 780  
 Qy 948 TEGCSLELLFQHPVQADTLTLMTSFFMESQVLPDTEILLEKESVHLGPLDTFCDIP 1007  
 Db 781 TEGCSLELLFQHPVQADTLTLMTSFFMESQVLPDTEILLEKESVHLGPLDTFCDIP 840  
 Qy 1008 TIKLAVDGKVSQVYTFPDERIEIDALLTSQHP5PLCSGCRPVRYQVLRDPPASGLPV 1067  
 Db 841 TIKLAVDGKVSQVYTFPDERIEIDALLTSQHP5PLCSGCRPVRYQVLRDPPASGLPV 900  
 Qy 1068 VVTHSRKFTDVEVTPGQWYOVYLAAG5ELGEASPPLNHIGAPYCGDGKX5BRIGEE 1127  
 Db 901 VVTHSRKFTDVEVTPGQWYOVYLAAG5ELGEASPPLNHIGAPYCGDGKX5BRIGEE 960  
 Qy 1128 CDDGDLVSDGQSKVCELEBGFNCVGP5LCVYB5DGI5CEP5RT5YVDCGIYTPKGY 1187  
 Db 961 CDDGDLVSDGQSKVCELEBGFNCVGP5LCVYB5DGI5CEP5RT5YVDCGIYTPKGY 1020

Qy	1188	DDQAAATAYSHENKKKCPVSLVYNGEPSSLCTSYHNDLPNHRLLTGMFPCVASENETOD	1247
Db	1021	LDQNAATRAYSSHEKKKCPVSLVYNGEPSSLCTSYHNDLPNHRLLTGMFPCVASENETOD	1080
Qy	1248	DRSEQPESSLKKEDEVMILKVCENRPGEARAIFILTTDGLVPGEHOOPTVLLYLTDRGSS	1307
Db	1081	DRSQQPESSLKKEDEVMILKVCENRPGEARAIFILTTDGLVPGEHOOPTVLLYLTDRGSS	1140
Qy	1308	NHSLGTYGLSCQNNPLIINTVTHQNVLPFHTTSSVILNFSRRVGI SAVALRTSSRIGLSA	1367
Db	1141	NHSLGTYGLSCQNNPLIINTVTHQNVLPFHTTSSVILNFSRRVGI SAVALRTSSRIGLSA	1200
Qy	1368	PSNCTISEDEGNNHQSGSCIRHPCGKODSCPSLLLDHADVNVCTSIGPGLMKCAITTCORGF	1427
Db	1201	PSNCTISEDEGNNHQSGSCIRHPCGKODSCPSLLLDHADVNVCTSIGPGLMKCAITTCORGF	1260
Qy	1428	ALQASGGYIRPMQKEILLTCSGSHMDQNVSLPYDCGVPDPSPSLVNYANFSCSGTKFKL	1487
Db	1261	ALQASGGYIRPMQKEILLTCSGSHMDQNVSLPYDCGVPDPSPSLVNYANFSCSGTKFKL	1320
Qy	1488	RCSISCVPPAKLQGLSPWLTCLDEGLMSLPEVYCKLECDAPPIILNANLLPHCLQDNHD	1547
Db	1321	RCSISCVPPAKLQGLSPWLTCLDEGLMSLPEVYCKLECDAPPIILNANLLPHCLQDNHD	1380
Qy	1548	VGTICKYCKRGYVYAESAEGKVRNKLKIQCLBEGIWEQSGCIPVCEPPPEEGMYE	1607
Db	1381	VGTICKYCKRGYVYAESAEGKVRNKLKIQCLBEGIWEQSGCIPVCEPPPEEGMYE	1440
Qy	1608	CTNGFSLDSDQVLNCGNERKPLPILCTNEGIMTQEFKLCENLQGECPPPPELSNVEYKC	1667
Db	1441	CTNGFSLDSDQVLNCGNERKPLPILCTNEGIMTQEFKLCENLQGECPPPPELSNVEYKC	1500
Qy	1668	EOGIGIGAVCSPLCYIPSPSDPVMLEPNTTADTLEHMMEPVYVQSI VCTGRQWHPDPVLV	1727
Db	1501	EOGIGIGAVCSPLCYIPSPSDPVMLEPNTTADTLEHMMEPVYVQSI VCTGRQWHPDPVLV	1560
Qy	1728	HCIQSCBEPFADGWCDDTINNRAVCHYDGDGDCSSSTLSSSKYVIPFAADCDLDECTCRDPKA	1787
Db	1561	HCIQSCBEPFADGWCDDTINNRAVCHYDGDGDCSSSTLSSSKYVIPFAADCDLDECTCRDPKA	1620
Qy	1788	EENQ 1791	
Db	1621	EENQ 1624	
RESULT 6			
ABB84494			
ID	ABB84494	standard; Protein; 1385 AA.	
XX	ABB84494;		
XX			
XX	24-DEC-2002	(first entry)	
XX			
DE	Human pregnancy associated protein PAPP-Bc SEQ ID 16.		
XX			
KW	PAPP-E; human; pregnancy associated plasma protein E; abortive; contraceptive; gene therapy; vaccine; pregnancy; antenatal; diagnosis;		
KW	dysgenetic pregnancy.		
XX			
OS	Homo sapiens.		
XX			
XX	US2002102252-A1.		
XX			
PD	01-AUG-2002.		
XX			
PF	06-APR-2001; 2001JUS-0827998.		
XX			
PR	26-MAY-2000; 2000JUS-207456P.		
XX			
XX	(GUTY/) GU Y.		
XX	PA	(SHAN/) SHANNON M E.	
XX			

[illegible]

Qy	661	SPKAYMSVKEALQJNUSHPLNITYASSVRBDLQAATWPMDOXAVHILGGIYVSPA	720
Db	307	-----GGIYVSPA	314
Qy	721	YYGMPGHTDTHIHEVGHVLGYHYVFKGVSERESCNDPCKETVPMSMETDLCADTAPTPKS	780
Db	315	YYGMPGHTDTHIHEVGHVLGYHYVFKGVSERESCNDPCKETVPMSMETDLCADTAPTPKS	374
Qy	781	ELCEPEPTSDTCGFTREPGAPFTNYMSYTDNCTDNFTNQVARMHCYLDLVYQOMTES	840
Db	375	ELCEPEPTSDTCGFTREPGAPFTNYMSYTDNCTDNFTNQVARMHCYLDLVYQOMTES	434
Qy	841	RKPPPIPIPPNAVITQTNKSLTIHMLPRISGVYVDRASSLGCAGCTEDGTPQYVYHTASR	900
Db	435	RKPPPIPIPPNAVITQTNKSLTIHMLPRISGVYVDRASSLGCAGCTEDGTPQYVYHTASR	494
Qy	901	RVCSSSGWTEPEEAVGPPDVOPCEPSLOQASPEVHLVHMNMVPCPREGSLETLFQHP	960
Db	495	RVCSSSGWTEPEEAVGPPDVOPCEPSLOQASPEVHLVHMNMVPCPREGSLETLFQHP	554
Qy	961	VOADTLTLMVTSFPMESQVLFDEILLENKESVHLGLDFFCDIPLTIKLVHGKXSGV	1020
Db	555	VOADTLTLMVTSFPMESQVLFDEILLENKESVHLGLDFFCDIPLTIKLVHGKXSGV	614
Qy	1021	KVYTFDERIEIDALLTSQPHSPUCSGCRPYRYOVLRDPPASGLPVVYVTHSHKFTDVE	1080
Db	615	KVYTFDERIEIDALLTSQPHSPUCSGCRPYRYOVLRDPPASGLPVVYVTHSHKFTDVE	674
Qy	1081	VTPOQMVOYVLAEGAGELGESAPLNIHIGAPYCGDGKVSERLGEBCDDODIVSGDCS	1140
Db	675	VTPOQMVOYVLAEGAGELGESAPLNIHIGAPYCGDGKVSERLGEBCDDODIVSGDCS	734
Qy	1141	KVCELEBGFNCVGBEPLCYMWEGBGICBEPFERKTSIVDCGITYPBGYLDQWATPAYSHE	1200
Db	735	KVCELEBGFNCVGBEPLCYMWEGBGICBEPFERKTSIVDCGITYPBGYLDQWATPAYSHE	794
Qy	1201	DKKKCPVSLVTEGBHSLICTSYHNPDLPHNRPLTGMFPCVASENETODDRSQPREGSLKE	1266
Db	795	DKKKCPVSLVTEGBHSLICTSYHNPDLPHNRPLTGMFPCVASENETODDRSQPREGSLKE	854
Qy	1261	DEWVLKVCENRPGERAFIFLITDGLVRGBHQOTVTLYLTVDRGSHSLGTYGLSCQH	1320
Db	855	DEWVLKVCENRPGERAFIFLITDGLVRGBHQOTVTLYLTVDRGSHSLGTYGLSCQH	914
Qy	1321	NPLIINTVHONVLFPHITTSVILNPSRPVGISAVALTSRIGLSAPNSCISEBEGNH	1380
Db	915	NPLIINTVHONVLFPHITTSVILNPSRPVGISAVALTSRIGLSAPNSCISEBEGNH	974
Qy	1381	QGOGSCIHRCCKODSCPSLILDHADVNCISIGPILMCAITCQGRFALOASSQGYIIPM	1440
Db	975	QGOGSCIHRCCKODSCPSLILDHADVNCISIGPILMCAITCQGRFALOASSQGYIIPM	1034
Qy	1441	QKEILLITCSSGHMDONVSLPVDGVPBPSLVNANFSCSGTKFLKRCISICVPBPAKQ	1500
Db	1035	QKEILLITCSSGHMDONVSLPVDGVPBPSLVNANFSCSGTKFLKRCISICVPBPAKQ	1094
Qy	1501	GLSPMLTCLBGLMSLPEVYCKLEBDAPRIILNANLPLPHCLQDNHDVGTICXKECKRGY	1566
Db	1095	GLSPMLTCLBGLMSLPEVYCKLEBDAPRIILNANLPLPHCLQDNHDVGTICXKECKRGY	1154
Qy	1561	YVAESAEBKVNKLKLTQCLEGGIIEQSSCIPVCEPBPVPEGMYECTNPFSLDSQCVL	1620
Db	1155	YVAESAEBKVNKLKLTQCLEGGIIEQSSCIPVCEPBPVPEGMYECTNPFSLDSQCVL	1214
Qy	1621	NCNOBEREKLPLICTYEGJMTQBFKJCNULQGBCCPBPSELNSVEYKCEQGYIGIIVACSPL	1680
Db	1215	NCNOBEREKLPLICTYEGJMTQBFKJCNULQGBCCPBPSELNSVEYKCEQGYIGIIVACSPL	1274
Qy	1681	CVIPSPSDVMLPENTADTLEHMMBPVAVOSIVCIGRQOWMPDPLVNCISCEPFOADG	1744
Db	1275	CVIPSPSDVMLPENTADTLEHMMBPVAVOSIVCIGRQOWMPDPLVNCISCEPFOADG	1334

**RESULT 7**

ID ABG11138 standard; Protein; 1752 AA.

AC ABG11138;

DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #11129.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
Food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

PV WO200175067-A2.

PI 11-OCT-2001.

PR 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

FI Drmanac RT, Liu C, Tang YT;

DR WP1; 2001-639362/73.  
DR N-PSTB; AAS75325.

PT New isolated polynucleotide and encoded polypeptides, useful in  
diagnostics, forensics, gene mapping, identification of mutations  
responsible for genetic disorders or other traits and to assess  
biodiversity -  
  
Claim 20; SEQ ID NO 41497; 103bp; English.

The invention relates to isolated polynucleotide (I) and  
polypeptide (II) sequences. (I) is useful as hybridisation probes,  
polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
and gene mapping, and in recombinant production of (II). The  
polynucleotides are also used in diagnostics as expressed sequence tags  
for identifying expressed genes. (I) is useful in gene therapy techniques  
to restore normal activity of (II) or to treat disease states involving  
(II). (II) is useful for generating antibodies against it, detecting or  
quantitating a polypeptide in tissue, as molecular weight markers and as  
a food supplement. (II) and its binding partners are useful in medical  
imaging of sites expressing (II). (I) and (II) are useful for treating  
disorders involving aberrant protein expression or biological activity.  
The polypeptide and polynucleotide sequences have applications in  
diagnostics, forensics, gene mapping, identification of mutations  
responsible for genetic disorders or other traits to assess biodiversity  
and to produce other types of data and products dependent on DNA and  
amino acid sequences. ABG00010-ABG30377 represent novel human  
diagnostic amino acid sequences of the invention.  
Note: The sequence data for this patent did not appear in the printed  
specification, but was obtained in electronic format directly from WIPO  
at ftp.wipo.int/pub/published\_pct\_sequences.

Sequence 1752 AA;

Query Match 39.7%; Score 3909.5; DB 22; Length 1752;  
Best Local Similarity 45.8%; Pred. No. 3.3e-254;  
Matches 717; Conservative 294; Mismatches 502; Indels 53; Gaps 25

249 REATFNSSQGLP--ILYFGRRRL-LRPVLAEIRFAETVAWKPKSGGNPPAI 305



Db 153 REARGATEEPPSPRRALYFSGRGEQLRYLRADL--ELPRDAFTLLQWMLRAEGORSPAVI 210  
 Qy 306 AGVEDNCSTHTVSDGMALGIRSGDKGRDARFPFSLCTDRVKATLILSHSRQPGWMT 365  
 Db 211 TGLYDKCSYISRDGKMWVGHITISDQNDKDPFSLKTDKARQVTTINARSLYPGWV 270  
 Qy 366 HVAATYDGRNHALYVDGTQVAVSLDQSGPLNSPFMACSRSLILGDSSEDEGHYRGHLGT 425  
 Db 271 YLAATYDQGFVKLYVNGAQVATSGEQVGIFSPILQKCKVLMG--SALHNHNRGYIEH 328  
 Qy 426 LVFVSTALPQSHFQHSQHSQSGEEBADLVLTASFEPPVNTWVPREDKTYRLEY--LQG 483  
 Db 329 FSLKVKATQREILISDMTHGAHTALPQLLQEMNDVYKHAMSPKDDSSKVEFSNANG 388  
 Qy 484 FEPEBELSPLOPPLCGQTVCDNVELISQYNGWYPLRGEKVIROYVNICDEDEGINPVS 543  
 Db 389 FLUD---TSLRPLCGQLCDNTEVILASYNQLSFRQPKVRRVNVLYEDDHKNPVT 444  
 Qy 544 EEOIRLOHEALNEAFSRYNISQVHVNSTLRHRVVLVNCSEPSKIGNDCEPECHP 603  
 Db 445 REQYDFQHQLAEAFKQYINISWELDVLEVSNSLRRLILANCDISKIDENCDECHNT 504  
 Qy 604 LTGIDGDCR--LQGRCSYWNRRDGLCHVECNMMLNDFDGDCCDQVADVKTCFDPDP 662  
 Db 505 LTGHGDCGRHLRHPAFVKQHNVCMDMCNVERFNFQGECDPEITNVYQTCFDPDP 564  
 Qy 663 KRAYSVKELKEALQNLNTHFLNTYFASVREDLGAATWPMDDAVTHLGLVILSPAY 722  
 Db 565 HRAVLVDNELKNILLDGSTHLNITFAKSSSEBLGAVTWPMDEKALMHLGLVILNPSFY 624  
 Qy 723 GMPGHTDMIHVEGVHVLGLYHVFKGVSEESCNPDCKETVPSMETGDLCACTAPPKSEL 782  
 Db 625 GMPGHTDMIHVEIGSLGLYHVFPRGISLQSDPCMETEFSFETGDLCACTNAPKHS 684  
 Qy 783 CREPEPTSDTCGTFRRPAPRTYMSYTDNCTNFTPNQYARHNCYIDLTYQOWTSRK 842  
 Db 685 CGDGPFGNDTCGFHSFFMTPTNNFMSYADDDCTDSTFTNOYARHNCYIDLTYQOWTSRK 744  
 Qy 843 PTPPIPPMVIQONTKSLTIHMLPISGVVYDRASGLSGACTEDGTROYVHTASSRRV 902  
 Db 745 PAVYALAOVAGHTTDSVTLKMPRIDGHFERELGSGHCLLEBRILLVOYASNNSSMP 804  
 Qy 903 CDSGGYTPREAVGPPVDQCEPSLQAMSPPEVLHNNMTVPCP--TEGCSLELLFQHPV 961  
 Db 805 CSBPGHMSPRABEGHPDVEQCKSVRTWSPNSANPHTPVACBEPGQVLEFLYPL 864  
 Qy 962 QADTLTLMT--SEPMSSQVLPFTTEILLENKESVHLGRLDTCFDPITLKL--HYDGKVS 1018  
 Db 865 VPESLTIWTVFVSTDMDSGAVNDIKLAVSGKNSISLQPOVWFCVPLTIRLMDVGEERY 924  
 Qy 1019 GVKYVTPPERLEIDAALLTSQPHSLGSGCRPVRYOVLRRDPRPAGLRFVVVTHSHRFTD 1078  
 Db 925 GIQYITLDEHLEIDAMLSTADTPLCLQCKPLKVVDRDPLQMDVASIL--HLNRKFPD 983  
 Qy 1079 VEAVPGQWYQYOVLAAGELGEASPPRLNHTIGAPYCGDKVSEKLEECDDGLVSGDG 1138  
 Db 984 MDLNGSYQYQWVITISTESEPSPATYIHHGRYCGDDGIIQDQGGQCDMDMKINDG 1043  
 Qy 1139 CSKVCLEBEGNCGEBSLCYMBGDI:CEPERKTSIVDCGITYPKYLDQMATRAYSS 1198  
 Db 1044 CSLFCRQSVFNCIDEPSCYFHDDGVCSEFEQKTSIKDGGVYVPPQGFLOQMASNVS 1103  
 Qy 1199 HEDKKKCVSLVTGER--HSLICTSYHPLPMHRPLTGFPCVASENENRQDORSBQPSL 1257  
 Db 1104 HQD--QCCGQWYIIGQPAASQCRKTVIDLSEGISQAHAYPCTISTPYSQ----- 1151  
 Qy 1258 KKEDEWLVKVCFNPRGEARAFIFLTTGLVGEHQCTVTLYLTVDVGSNHSLQYGLS 1317  
 Db 1152 LAQTTFWLRAYFSGQWMAVAIVYHVTDTGYGQDKQKTIISQVLLDTDQSHDGLHLYLS 1211  
 Qy 1318 COHNPILINVTHTQVLFHHTTSVLANTSSPRVGISAVALTSSRIGLSADPNCISDEG 1377  
 Db 1212 CRNPPLIFVVDHLSQPFYHSAVAVSFSSPLVAISGVALRSFDFPVTLSSC--QRGET 1270

Qy 1378 QNHQGSICRRPGCKODSCPSLLLDHADVNVCTSI-----GPGIMKCAITCORGFALQASS 1433  
 Db 1271 YSPAGSCVAFACEKD--CELAIVENAS--LNCSSSDRYHG---AQCTVSCRTGYVLQIR 1325  
 Qy 1434 GQYIRPMQ--KEILLTSSGHDQNVSCLPVDGVPDPSPSLVNYANFSCSEGTFLKRCST 1491  
 Db 1326 DDELIKSQTPSVYVTCTEGKMNKQVACEPVDCSIPIHQVYAAASFSCPGETFGSQCSF 1385  
 Qy 1492 SCVPALQGLSPMLTLEBQGLMSLPREVYKLECDAPILLNALLPHLQDNHVDGTI 1551  
 Db 1386 QCRHPADLKNNLSLITMEDGLMSFPFALCEMLCLAPVPVADLQARRENHGRKGSF 1445  
 Qy 1552 CKYCEKGYVVAESAEKVNKLLKIQLEGLMEQSCIPVYCEPPVFEQWECTNG 1611  
 Db 1446 CKYCKCKGYHVGSSR--KSJKRAFKYQCTQDGSQWQBACVPCYCDPPPPFHLGYCTNG 1504  
 Qy 1612 FSLDSQCVLNC-----NQEREKLPILCTKEGLWTOEFKLCENLQGECPPEPSLNS--VEY 1665  
 Db 1505 FQFNSCEKICEDSDASQGLGSNVYHCKRQDGTWNGSFHYVQEQMGQC--SVPNELNSMLKL 1563  
 Qy 1666 KCEGYYGIGAVCSPLCYIPSPDPMLEPNTADTLLEHMPAVYQSIYCTGRQMHDPV 1725  
 Db 1564 QCPDGYAIGSECATSCIDHNSISILPMNVTVRDIPEMLNPTVRERVCTAGLKYEPHA 1623  
 Qy 1726 LVHCTQSCPPQADGMCCTINNRAVCHYDGDCCSSTLSSKVIYFPAADCLD--ECTGRD 1784  
 Db 1624 LIHCVKCEBPMGNDYCDALINNAFCYVDGDCCTSTVTKTKVTPFMSCDLQDCACARD 1683  
 Qy 1785 PKAEN 1790  
 Db 1684 PQGPRN 1689

RESULT 8  
 AAU32498  
 ID AAU32498 standard; Protein; 1603 AA.  
 XX AC AAU32498;  
 XX DT 18-DEC-2001 (first entry)  
 XX DE Novel human secreted protein #2989.  
 XX KW Human: vaccination; gene therapy; nutritional supplement;  
 KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;  
 KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.  
 XX OS Homo sapiens.  
 XX PN WO200179449-A2.  
 XX PD 25-OCT-2001.  
 XX PF 16-APR-2001; 2001WO-US08656.  
 XX PR 18-APR-2000; 2000US-0552929.  
 XX PR 26-JAN-2001; 2001US-0770160.  
 XX PA (HYSE-) HYSEQ INC.  
 XX PI Tang YT, Liu C, Drmanac RT;  
 XX DR WPI; 2001-611725/70.  
 XX PT Nucleic acids encoding a range of human polypeptides, useful in genetic  
 XX vaccination, testing and therapy -  
 XX PS Claim 20; Page 624; 765pp; English.  
 CC The invention relates to novel human secreted polypeptides. The  
 CC polypeptides and antibodies to the polypeptides are useful for  
 CC determining the presence of or predisposition to a disease associated



CC with altered levels of polypeptide. The polypeptides are also useful for  
 CC identifying agents (agonists and antagonists) that bind to them. Cells  
 CC expressing the proteins are useful for identifying a therapeutic agent  
 CC for use in treatment of a pathology related to aberrant expression or  
 CC physiological interactions of the polypeptide. Vectors comprising  
 CC the nucleic acids encoding the polypeptides and cells genetically  
 CC engineered to express them are also useful for producing the proteins.  
 CC The proteins are useful in genetic vaccination, testing and  
 CC therapy, and can be used as nutritional supplements. They may be used to  
 CC increase stem cell proliferation; to regulate haematopoiesis; and in  
 CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;  
 CC immune suppression and/or stimulation; as anti-inflammatory agents; and  
 CC in treatment of leukemias. AAU29510-AAU3304 represent the amino acid  
 CC sequences of novel human secreted proteins of the invention.

XX Sequence 1603 AA;

Query Match 38.7%; Score 3810; DB 22; Length 1603;  
 Best Local Similarity 45.0%; Pred. No. 1.5e-247;  
 Matches 704; Conservative 294; Mismatches 490; Indels 78; Gaps 27;

QY 249 REAFTHSQVLP--LYFSGRRL--LREVLAEIREFATYEAWKPEGGONPAII 305  
 DB 80 REAGATEEESPFPARALYFSGRGQLRVLRADL--ELPRDAFTQLVLRAGGGRSPAVI 137  
 QY 306 AGVDNCSHTVSDKGMALGIRSGDKRDARFFSLCTDRKATILISHRYOPGTWT 365  
 DB 138 TGLYDKCSYIRSDGKGVGIGHTISDQDKDPRYFSLKTDPAKQVTTINARSLIPGMV 197  
 QY 366 HVAATYDGRHMAIYVDTQVASSLDQGPLNSPFMAICRSLLGDSSEDEGHYRGLGT 425  
 DB 198 YLAATYDGFQFKLYVNGAQVATSGEYVGIFFSLQCKVLMG--SALNHNRYGYLEH 255  
 QY 426 LVFSTALPQSHFQSSGHSGBEARDLVLTASEPNTMEVPRDEKRYLVE--LQG 483  
 DB 256 FSLKLVKATQREILSDMETHGAHTALPQLLDENMDVKNAMSPKQSSKVEFSNAG 315  
 QY 484 FEPEELISPLQPCQGVCDNELLISQVNGYMPRLGEKIRYQVNNICDEGLNPIVS 543  
 DB 316 FLUD-----TSLPEPLCGOTLCDNTEVLASVYQSSFRPKYKRVVNLVYDDHKNPVYT 371  
 QY 544 EEQIRLOHEALNEAFSRYNISWOLSVHQNSTLRHRVVLVNCESKIGNDHCDPECHP 603  
 DB 372 REQVDPQHQAFAFKQYNISWELDVLEVNSSLRRRLILANCDISKIDENCDPECHT 431  
 QY 604 LTGIDGDCR-LQGRCSYANRBDGLCHVECNMLNDPFDGCCPOVADVKTCTPDPSP 662  
 DB 432 LTGHDGDCRHLRPAFAFKQHNQVCDMDCYERFNPFGGECDBEITNVYQTCFDPSP 491  
 QY 663 KRAYASYELKEALQNLNTHFLNIFYASVREDLAGAATWPMDOAVTHLGIYLSPAY 722  
 DB 492 HRAFLVDNELKILNLDGSTHLNIFAKSSEELAGVATWPMDEALNHLGIVLNPFY 551  
 QY 723 GMPGHTDIMEVGHVGLVHVFQGVSEESCNPKETVPSMETGDLCAOTAPPKSEL 782  
 DB 552 GMPGHTDIMEHIGHSIGLVHVFPGISISIGSDPC-----GD----- 589  
 QY 783 CREPEPISDTGCFTRFPCAPFTNYSYTDNCTDNFTPNQVARNHCYIDLVOQWTSRK 842  
 DB 590 ---GPGNDTCGFHSFNTPNNFMYSYADDCDTSFTPNQVARNHCYIDLVOQWTSRK 846  
 QY 843 PTPRIPMVIIGOTNKSILTIHMLPRISGVVYDRASGSLCGACTEDGTROYVHTASSRV 902  
 DB 647 PAPVALAPQVAGHTTDSVTELEMPRIDGHFERELIGSCHULCBORILVQVANSASMP 706  
 QY 903 CDSGGVTPBEAVGPVPVDPQCEPSLQAMSEVHLVHNMATVPGP-TEGSLTELLFQHPV 961  
 DB 707 CSPSGHMSPREAEGHPDVEGCKSVRTWSPNSAVNPHTPVACBEPGCGYLELEFLYPL 766  
 QY 962 QADTLTLMTV--SFEMSSQVLFPTTEILLKESVHLCPLDTPCDIPLTILV-HYDGKVS 1018  
 DB 767 VPESLITWIVTVSITDWDSSGAVNDIKLAVSGKNISLQPVQVFCVPLTIRLMDVGEVY 826

QY 1019 GVKVYTFDERIEIDALLTSQPHSLCSGRPVRYQVLRDPFASGLPVVWTHSHRKPTD 1078  
 DB 827 GIQVYTDENHEIDLAAMLTSADPTLCTQCKPKYKVRDPLQMDVASIL-HLNKKEVD 885  
 QY 1079 VEYTPGQMYQVLAELAGSELGEMASPRPLNHIKAPRYGDKGVSRLEBECDDGLVSGDG 1138  
 DB 886 MDNLGSGVYQVWVYITIGTESBSPVAVTYIHGKYGCGDGIQKDGEOCCDNKINGDG 945  
 QY 1139 CSKYCELEEGNCGEBSLCAWBGDDICPEPERKTSIVOCGITYPGVLDQMATRAYSS 1198  
 DB 946 CSLEFRQEVSNCLDEBSRCTFHGDGVCEFEQKTSIKDQGVYTPQGFLLQMSMASVS 1005  
 QY 1199 HEDKKKCPVSLVTEP-HSLICTSYHBDLPNHRLELTGMPFVASENTEODDRSQPEBSL 1257  
 DB 1006 HQD-QQCPGMVYISQPAASQVCRKTVIDLSEGISQNHMYPTTISYPSQ----- 1053  
 QY 1258 KKEDEWLVKVCENRPGARALFPLTDTGLVGEHQPTVYLTVDYRGSNHSIGTYGLS 1317  
 DB 1054 LAQTFPWLRAVYFSQPMVAALVYVLDGYGQKQKETSIVQLDTRKQSHDGLVHLS 1113  
 QY 1318 COHNPILINTYHQNVLFEHHTTSVLLNPSRPVGISVALRTSRIGLSAPNSCISEDEG 1377  
 DB 1114 CRNNPLIIPVNHDSQPPYHSQAVRVSFSSPLVALISGVALRSPDNPPVTLSSC-QRGET 1172  
 QY 1378 QNHQSGSCIRPSCGKODSCPSLLDHDVNVCTSI---GPGLMKCAITCORFALQASS 1433  
 DB 1173 YSPAEQSCVHPACGKTD-CBLAVENNS-LNCSSDHYHG---AQCTVSCRTGYVLQIRR 1227  
 QY 1434 QGYTRPMQ--KEILLTSSGHWDQVNSCLPYDCVPPPSLVYVNVANFSCSECTKFLKRCST 1491  
 DB 1228 DDELKIQGTGSPVTVTCTEGKMNQVACEPVDCSI PHNHQVYASFSCEPBGTTGSGQCSF 1287  
 QY 1492 SCVPRAKLGSLPMLTCLDGLWSLREVYCKLECDAPRIILNALLPHCLQDNHNDCTI 1551  
 DB 1288 QCRHPADLKNNLSLTMEDGLMSFREALCELMCLAPPVNPADLQTRARENHKGVSF 1347  
 QY 1552 CRYECKPGYVAESAEBKVENKLLKIQLEGGIWEQSCIPVVCPEPPVPEGVYECTNG 1611  
 DB 1348 CKYCKCKGYHVPSSR-KSKKRAKTOQTQDGSQNBGACVPTDPPRPKHGLYQCTNG 1406  
 QY 1612 FSLDSQVLNC-----NOERBKPLICTKGLWTOEFLCENIQGCEPPPSBLNS-VKY 1665  
 DB 1407 FQFNSECRICKCEDSDAQQGASVNIHCRKDGTVNGSFHVQCEMGQC-SVNEINLSNKL 1465  
 QY 1666 KCEBGYIGAVCSPLCYIIPSPDPMLEBNTADTLERHMEVVKQSIYVCTGRROMHPRV 1725  
 DB 1466 QCPBGVAIGSCATSCLDHNSESIILPMNVTVRDI PHMLPTRYERVYCTAGLKYPRPA 1525  
 QY 1726 LVHICGCEPFOADGMCDTINRAYCHYDGDCCSSTLSKRVLPFADCDLD-ECTGRD 1784  
 DB 1526 LIFCVKGCPEPMDNYCDALNNRAFCNYDGDCCSTIVTKYKVTFFPMSCDLQGDCCARD 1585  
 QY 1785 PKAEN 1790  
 DB 1586 PQAQEH 1591

RESULT 9  
 AAM24060  
 ID AAM24060 standard; Protein, 317 AA.  
 AC AAM24060;  
 DT 12-OCT-2001 (first entry)  
 DE Human EST encoded protein SEQ ID NO: 1585.  
 DE Human, sheep, pig, cow, fruit fly, yeast, hamster, macaque, horse;  
 DE tomato, monkey, dog, sea urchin, expressed sequence tag; EST;  
 DE diagnostics; forensic test; gene mapping; genetic disorder;  
 DE biodiversity; gene therapy; nutrition.  
 OS Homo sapiens.

XX WO200154477-A2.  
XX  
XX 02-AUG-2001.  
XX  
XX 25-JAN-2001; 2001WO-US02687.  
XX  
XX 25-JAN-2000; 2000US-0491404.  
XX 17-JUL-2000; 2000US-0617746.  
XX 03-AUG-2000; 2000US-061451.  
XX 15-SEP-2000; 2000US-0663870.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;  
XX Cao Y, Drmanac RA, Zhang J, Werhman T;  
XX  
XX WPI; 2001-476164/51.  
XX N-PSDB; AAH98719.  
XX  
XX Isolated polypeptide for treatment of diseases, diagnostics, raising  
XX antibodies and research use -  
XX  
XX Claim 20; Page 1084; 1275pp; English.  
XX  
XX The present invention provides the protein and coding sequences of novel  
XX proteins from a variety of organisms, including human, dog, cat, horse,  
XX cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea  
XX urchin and tomato. These were derived from expressed sequence tags (ESTs)  
XX from the organism of interest. They can be used in diagnostics,  
XX forensics, gene mapping, identification of mutations, to assess  
XX biodiversity and for nutritional purposes. The present sequence is a  
XX protein of the invention.  
XX  
XX Sequence 317 AA;  
SQ  
Query Match 16.5%; Score 1629; DB 22; Length 317;  
Best Local Similarity 100.0%; Pred. No. 1,7e-101;  
Matches 307; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MMCKIIRISLAIAGMALCSANSELGWTTRKSLIVERHNLQVLLGRCWLGAQVRRPR 60  
DB 1 MMCKIIRISLAIAGMALCSANSELGWTTRKSLIVERHNLQVLLGRCWLGAQVRRPR 60  
QY 61 ASPQHLFGVPSRAGNYLRYPVGEORIHHTGSKPTBEGNAVSLVPPDLTENPAGLRG 120  
DB 61 ASPQHLFGVPSRAGNYLRYPVGEORIHHTGSKPTBEGNAVSLVPPDLTENPAGLRG 120  
QY 121 AVEBPAPWVGDSPIGSELGDDDAYLGNORSKESLGEAGIOKGSMAATTTTAIFTTL 180  
DB 121 AVEBPAPWVGDSPIGSELGDDDAYLGNORSKESLGEAGIOKGSMAATTTTAIFTTL 180  
QY 181 NEPKETORGMWAKSRORQVWKRRAEDQGDGSGISHFQWPKSLKRVKSPPEESN 240  
DB 181 NEPKETORGMWAKSRORQVWKRRAEDQGDGSGISHFQWPKSLKRVKSPPEESN 240  
QY 241 ONGEGSYREAFETNSQVGLPIFYSGRRERLLRPETLARIPEAFVEMVWPEGON 300  
DB 241 ONGEGSYREAFETNSQVGLPIFYSGRRERLLRPETLARIPEAFVEMVWPEGON 300  
QY 301 NPATIAAG 307  
DB 301 NPATIAAG 307  
RESULT 10  
ABG48354  
ID ABG48354 standard; Peptide; 192 AA.  
XX  
XX ABG48354;  
XX  
XX 25-FEB-2003 (first entry)  
XX

DE Human liver peptide, SEQ ID No 27002.  
XX  
XX Human; liver; cirrhosis; hyperlipoproteinemia; hyperlipidaemia;  
XX hypercholesterolaemia; coronary heart disease.  
XX  
XX Homo sapiens.  
XX  
XX WO200157273-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX 30-JAN-2001; 2001WO-US00664.  
XX  
XX 04-FEB-2000; 2000US-0180312.  
XX 26-MAY-2000; 2000US-0207456.  
XX 30-JUN-2000; 2000US-0608408.  
XX 03-AUG-2000; 2000US-0632366.  
XX 21-SEP-2000; 2000US-0234687.  
XX 27-SEP-2000; 2000US-0236359.  
XX 04-OCT-2000; 2000GB-0024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2001-468898/53.  
XX  
XX Human genome-derived single exon nucleic acid probes useful for  
XX analyzing gene expression in human adult liver -  
XX  
XX Claim 27; SEQ ID No 27002; 658pp; English.  
XX  
XX The invention relates to a single exon nucleic acid probe (SENP) (I) for  
XX measuring human gene expression in a sample derived from human adult  
XX liver, comprising one of 13109 defined nucleotide sequences given in the  
XX specification (or complements/ fragments). The probe hybridizes at high  
XX stringency to a nucleic acid molecule expressed in the human adult  
XX liver. (I) may be used for predicting, measuring and displaying gene  
XX expression in samples derived from human adult liver. The genes  
XX identified may be involved in genetic liver diseases such as cirrhosis,  
XX hyperlipoproteinemia, hyperlipidaemia and hypercholesterolaemia which  
XX is associated with coronary heart disease. ABG47348-ABG59930 represent  
XX human liver single exon encoded peptides of the invention.  
XX Note: The sequence information for this patent does not appear in the  
XX printed specification but was obtained in electronic format directly  
XX from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 192 AA;  
SQ  
Query Match 11.0%; Score 1086; DB 22; Length 192;  
Best Local Similarity 100.0%; Pred. No. 3.4e-65;  
Matches 192; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 472 DEKYPRLLEVLOGEPEPEELISPLQPLCGQTVCDNVELISQNGWPLRGEKYIRYQVN 531  
DB 1 DEKYPRLLEVLOGEPEPEELISPLQPLCGQTVCDNVELISQNGWPLRGEKYIRYQVN 60  
QY 532 ICDDEGLNPVISEQQLQHEALNEAFSRYNISQQLSVHQVHNSTLRAHRYVLVNCESKI 591  
DB 61 ICDDEGLNPVISEQQLQHEALNEAFSRYNISQQLSVHQVHNSTLRAHRYVLVNCESKI 120  
QY 592 GNDHCDPECEHPLTGYGSDGCRLOGRCYSNRRDGLCHVECNMNLNDFDGDCCDDPOVAD 651  
DB 121 GNDHCDPECEHPLTGYGSDGCRLOGRCYSNRRDGLCHVECNMNLNDFDGDCCDDPOVAD 180  
QY 652 VRKTCFDDSPK 663  
DB 181 VRKTCFDDSPK 192  
RESULT 11  
ABB28331  
ID ABB28331 standard; Peptide; 192 AA.

```

XX ABB28311,
AC 01-FEB-2002 (first entry)
DT Human peptide #982 encoded by breast cell single exon nucleic acid probe.
DE Human; microarray; single exon probe; gene expression; breast;
XX disease; cancer.
XX Homo sapiens.
XX WO200157271-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US00662.
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-496933/54.
XX New spatially-addressable set of single exon nucleic acid probes,
PT useful for measuring gene expression in sample derived from human
PT breast, comprises number of single exon nucleic acid probes
XX Claim 27, SEQ ID NO 11299; 327pp + sequence listing; English.
XX The invention relates to a spatially-addressable set of single exon
XX nucleic acid probes for measuring gene expression in a sample derived
XX from human breast and BT 474 cells. The method involves contacting
XX the probes with a collection of detectably labelled nucleic acids
XX derived from mRNA of human breast, and then measuring the label
XX bound to each probe of the microarray. The probes are useful for
XX verifying the expression of regions of genomic DNA predicted to
XX encode proteins. They are useful for gene discovery, and for
XX determining predisposition and/or prognosing breast disease. Gene
XX expression analysis is useful for assessing the toxicity of chemical
XX agents on cells. The microarray of this invention presents a far greater
XX diversity of probes for measuring gene expression, with far less bias
XX than expressed sequence tag microarrays. The method is suitable for
XX rapid production of functional information from genomic sequence. The
XX present sequence is a peptide encoded by a single exon nucleic acid
XX probe of the invention.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ffp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 192 AA;
Query Match 11.0%; Score 1086; DB 22; Length 192;
Best Local Similarity 100.0%; Pred. No. 3.4e-65;
Matches 192; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 472 DEKXPRLEVLQGFPEPEILSPLOPLCGQTVCDNVELISQYNGWPLRGKVIYQVNN 531
DB 1 DEKXPRLEVLQGFPEPEILSPLOPLCGQTVCDNVELISQYNGWPLRGKVIYQVNN 60
QY 532 ICDDEGNLPVSEQIRLOHEALNEAFSRYNISWQSVHVNSTLHRVVLVNCESPSKI 591
DB 61 ICDDEGNLPVSEQIRLOHEALNEAFSRYNISWQSVHVNSTLHRVVLVNCESPSKI 120
QY 592 GNDHCDPECEHPLTGYDGDRCRLQGRCYSWNRBDGLCHVECNMMLNDFDGDCCDPQVAD 651

```

```

DB 121 GNDHCDPECEHPLTGYDGDRCRLQGRCYSWNRBDGLCHVECNMMLNDFDGDCCDPQVAD 180
QY 652 VRKTCFDPDSPK 663
DB 181 VRKTCFDPDSPK 192
RESULT 12
ABB33508
ID ABB33508 standard; Peptide, 192 AA.
XX ABB33508;
AC 04-FEB-2002 (first entry)
DT Peptide #1014 encoded by human foetal liver single exon probe.
DE Human; foetal liver; gene expression; single exon nucleic acid probe.
XX Homo sapiens.
XX WO200157277-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US00669.
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483447/52.
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver -
PT
XX Claim 27, SEQ ID NO 26143; 639pp + sequence listing; English.
XX The invention relates to a single exon nucleic acid probe for
XX measuring human gene expression in a sample derived from human foetal
XX liver. The single exon nucleic acid probes may be used for predicting,
XX measuring and displaying gene expression in samples derived from human
XX fetal liver. The present sequence is a peptide encoded by a single exon
XX nucleic acid probe of the invention.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ffp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 192 AA;
Query Match 11.0%; Score 1086; DB 22; Length 192;
Best Local Similarity 100.0%; Pred. No. 3.4e-65;
Matches 192; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 472 DEKXPRLEVLQGFPEPEILSPLOPLCGQTVCDNVELISQYNGWPLRGKVIYQVNN 531
DB 1 DEKXPRLEVLQGFPEPEILSPLOPLCGQTVCDNVELISQYNGWPLRGKVIYQVNN 60
QY 532 ICDDEGNLPVSEQIRLOHEALNEAFSRYNISWQSVHVNSTLHRVVLVNCESPSKI 591
DB 61 ICDDEGNLPVSEQIRLOHEALNEAFSRYNISWQSVHVNSTLHRVVLVNCESPSKI 120
QY 592 GNDHCDPECEHPLTGYDGDRCRLQGRCYSWNRBDGLCHVECNMMLNDFDGDCCDPQVAD 651

```

Db 121 GNDHCDPECEHPLTGYDGGDCRLQGRCTSWNRBDGLCHVECNMNLNDFDGDCCDPQVAD 180  
Qy 652 VRKTCFDPDSPK 663  
Db 181 VRKTCFDPDSPK 192

RESULT 13  
ABBI8967  
ID ABBI8967 standard; Protein; 192 AA.  
XX  
AC ABBI8967;  
XX  
DT 23-JAN-2002 (first entry)  
XX  
DE Protein #966 encoded by probe for measuring heart cell gene expression.  
XX  
KW Human; gene expression; heart; microarray; vascular system;  
KM cardiovascular disease; hypertension; cardiac arrhythmia;  
XX congenital heart disease.  
XX  
OS Homo sapiens.  
XX  
PN WO200157274-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US00666.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-488899/53.  
XX  
DR WPI; 2001-488899/53.  
XX  
PT Single exon nucleic acid probes for analyzing gene expression in human  
PT hearts -  
XX  
PS Claim 15; SEQ ID No 20737; 530pp; English.  
XX  
XX The present invention relates to single exon nucleic acid probes for  
CC measuring human gene expression in a sample derived from human heart (see  
CC AB21535-ABA1305). The present sequence is a protein encoded by one such  
CC probe. The probes may be used for predicting, measuring and displaying  
CC gene expression in samples derived from the human heart via microarrays.  
CC By measuring gene expression, the probes are useful for predicting,  
CC diagnosing, grading, staging, monitoring and prognosing diseases of the  
CC human heart and vascular system e.g. cardiovascular disease.  
CC hypertension, cardiac arrhythmias and congenital heart disease.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX  
SQ Sequence 192 AA;  
Query Match 11.0%; Score 1086; DB 22; Length 192;  
Best Local Similarity 100.0%; Pred. No. 3,4e-65;  
Matches 192; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 472 DEKYPRLLEVQGFPEPEIISPLQPLCGQTVCDNVELISQYNGWPLRGEKVIROYVN 531  
Db 1 DEKYPRLLEVQGFPEPEIISPLQPLCGQTVCDNVELISQYNGWPLRGEKVIROYVN 60  
Qy 532 ICDDGGLNPVSEBQIRLOHEALNEAFSRYNISWQLSVHGVHNSTLRHRVVLVNCPEPSKI 591  
Db 1 ICDDGGLNPVSEBQIRLOHEALNEAFSRYNISWQLSVHGVHNSTLRHRVVLVNCPEPSKI 591

Db 61 ICDDGGLNPVSEBQIRLOHEALNEAFSRYNISWQLSVHGVHNSTLRHRVVLVNCPEPSKI 120  
Qy 592 GNDHCDPECEHPLTGYDGGDCRLQGRCTSWNRBDGLCHVECNMNLNDFDGDCCDPQVAD 651  
Db 121 GNDHCDPECEHPLTGYDGGDCRLQGRCTSWNRBDGLCHVECNMNLNDFDGDCCDPQVAD 180

Qy 652 VRKTCFDPDSPK 663  
Db 181 VRKTCFDPDSPK 192

RESULT 14.  
AAM54287  
ID AAM54287 standard; Protein; 192 AA.  
XX  
AC AAM54287;  
XX  
DT 05-NOV-2001 (first entry)  
XX  
DE Human brain expressed single exon probe encoded protein SEQ ID NO: 26392.  
XX  
KW Human; brain expressed exon; gene expression analysis; probe;  
KM microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
XX epilepsy; cancer.  
XX  
OS Homo sapiens.  
XX  
PN WO200157275-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US00667.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-483446/52.  
XX  
DR WPI; 2001-483446/52.  
XX  
PT Single exon nucleic acid probes for analyzing gene expression in human  
PT brains -  
XX  
PS Example 4; SEQ ID NO: 26392; 650pp + Sequence Listing; English.  
XX  
XX The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC brain. They can be used to measure gene expression in brain cell samples,  
CC which may enable the diagnosis and improved treatment of nervous system  
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
CC epilepsy and cancers. The present sequence is a protein encoded by one of  
CC the probes of the invention.  
XX  
XX  
SQ Sequence 192 AA;  
Query Match 11.0%; Score 1086; DB 22; Length 192;  
Best Local Similarity 100.0%; Pred. No. 3,4e-65;  
Matches 192; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 472 DEKYPRLLEVQGFPEPEIISPLQPLCGQTVCDNVELISQYNGWPLRGEKVIROYVN 531  
Db 1 DEKYPRLLEVQGFPEPEIISPLQPLCGQTVCDNVELISQYNGWPLRGEKVIROYVN 60  
Qy 532 ICDDGGLNPVSEBQIRLOHEALNEAFSRYNISWQLSVHGVHNSTLRHRVVLVNCPEPSKI 591  
Db 61 ICDDGGLNPVSEBQIRLOHEALNEAFSRYNISWQLSVHGVHNSTLRHRVVLVNCPEPSKI 120



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using SW model

Run on: January 2, 2004, 16:02:37 ; Search time 21 Seconds

(without alignments)  
4010.706 million cell updates/sec

Title: US-09-983-025a-2  
Perfect score: 9856  
Sequence: 1 MMCLIKIRISLAIAGWALC.....AADCDIDECTCRDPKAEENQ 1791

Scoring table: BLAST62  
Gapop 10.0, Gapext 0.5

Searched: 127863 seqs, 47026705 residues  
Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database: SwissProt\_41:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3916.5	39.7	1627	1 PAPA_HUMAN	Q13219 homo sapien
2	287.5	2.9	2039	1 CRI_HUMAN	P17927 homo sapien
3	249	2.5	830	1 LEM3_HUMAN	P16109 homo sapien
4	249	2.5	1033	1 CR2_HUMAN	P20023 homo sapien
5	247	2.5	612	1 LEM2_MOUSE	Q00690 mus musculu
6	238.5	2.4	1234	1 CFAH_MOUSE	P06909 mus musculu
7	230.5	2.3	610	1 LEM2_HUMAN	P16581 homo sapien
8	229	2.3	407	1 DAF1_MOUSE	Q61476 mus musculu
9	227.5	2.3	768	1 LEM3_RAT	P27113 ratu mus norv
10	225	2.3	551	1 LEM2_RABIT	P27113 ratu mus norv
11	225	2.3	1394	1 LEM3_HUMAN	P22064 homo sapien
12	223	2.3	390	1 DAF1_MOUSE	Q61476 mus musculu
13	221	2.2	381	1 DAF1_MOUSE	P08174 homo sapien
14	219	2.2	768	1 LEM3_MOUSE	Q01102 mus musculu
15	218.5	2.2	1025	1 CR2_MOUSE	P19070 mus musculu
16	218	2.2	1712	1 LEM1_RAT	Q00918 ratu mus norv
17	217	2.2	5376	1 ZAN_MOUSE	Q08879 mus musculu
18	215.5	2.2	1713	1 LEM3_MOUSE	Q08879 mus musculu
19	214	2.2	769	1 LEM3_SHEEP	Q08879 mus musculu
20	213.5	2.2	1231	1 CFAH_HUMAN	P08603 homo sapien
21	212	2.2	1696	1 PCK5_BRACL	Q09115 brachiosto
22	211	2.1	1389	1 LEM3_MOUSE	Q08879 mus musculu
23	203	2.1	2471	1 NTC2_RAT	Q09430 ratu mus norv
24	202	2.0	646	1 LEM3_BOVIN	P42201 bos taurus
25	201	2.0	3133	1 HMCT_BOVMO	P98092 bombyx mori
26	200	2.0	558	1 C4BP_RAT	Q63514 ratu mus norv
27	199.5	2.0	611	1 LEM2_CANFA	P33730 canis famil
28	198	2.0	2437	1 NTC1_BRARE	P46530 brachiosteo
29	197.5	2.0	610	1 C4BP_BOVIN	Q28065 bos taurus
30	196	2.0	1595	1 LEM3_HUMAN	Q14766 homo sapien
31	196	2.0	1485	1 LEM2_BOVIN	Q14766 bos taurus
32	196	2.0	1877	1 PCK5_MOUSE	Q04592 mus musculu
33	194	2.0	1906	1 AT20_MOUSE	P59511 mus musculu

34	193.5	2.0	597	1 C4BP_HUMAN	P04003 homo sapien
35	193.5	2.0	2871	1 FBN1_MOUSE	Q61554 mus musculu
36	193	2.0	668	1 F13H_MOUSE	Q07968 mus musculu
37	192.5	2.0	549	1 LEM2_RAT	P98105 ratu mus norv
38	192	1.9	2556	1 NTC1_HUMAN	P46531 homo sapien
39	191.5	1.9	340	1 DAF1_MOUSE	P49457 pongo pygma
40	189	1.9	685	1 CFAH_BOVIN	Q28065 bos taurus
41	189	1.9	1429	1 L112_CAEEL	P14585 caenorhabdi
42	189	1.9	3084	1 LEM1_MOUSE	P19137 mus musculu
43	188.5	1.9	2871	1 FBN1_BOVIN	P98133 bos taurus
44	188.5	1.9	3767	1 MUO3_CAEEL	P34576 caenorhabdi
45	187	1.9	484	1 LEM2_PIG	P98110 sus scrofa

## ALIGNMENTS

1026

RESULT 1  
PAPA\_HUMAN STANDARD: PRT; 1627 AA.  
ID Q13219; 008371; Q9UDX7;  
AC 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Pregnancy-associated plasma protein-A precursor (EC 3.4.24.-) (PAPP-A)  
DE (Insulin-like growth factor-dependent IGF binding protein-4 protease)  
DE (IGF-dependent IGFBP-4 protease) (IGFBP-4ase).  
GN PAPP.  
OS Homo sapiens (human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxId=9606;  
RN [1]  
RP SEQUENCE FROM N.A. AND INDUCTION.  
RC TISSUE=Placenta;  
RA MEDLINE=96203921; PubMed=8620868;  
RX Haaning J., Oxvig C., Overgaard M.T., Ebbesen P., Kristensen T.,  
RA Sottrup-Jensen L.;  
RT "Complete cDNA sequence of the preproform of human pregnancy-  
RT associated plasma protein-A. Evidence for expression in the brain and  
RT induction by cAMP."  
RT Eur. J. Biochem. 237:159-163(1996).  
RN [2]  
RP SEQUENCE OF 77-1627 FROM N.A. SEQUENCE OF 81-98; 117-126; 210-224;  
RP 466-485; 507-519; 576-593; 609-621; 718-736; 742-754; 1006-1017;  
RP 1259-1273; 1369-1374; 1389-1398; 1490-1509; 1524-1533 AND 1537-1544,  
RP VARIANT SER-944, AND TISSUE SPECIFICITY.  
RC TISSUE=Placenta, and Serum;  
RX MEDLINE=94146014; PubMed=7508748;  
RA Kristensen T., Oxvig C., Sand O., Moller N.P.H., Sottrup-Jensen L.;  
RT "Amino acid sequence of human pregnancy-associated plasma protein-A  
RT derived from cloned cDNA."  
RT Biochemistry 33:1592-1598(1994).  
RN [3]  
RP SEQUENCE OF 81-89; 117-126; 210-224; 460-485; 507-519; 576-593;  
RP 718-736; 742-754; 1259-1273; 1369-1374; 1490-1509; 1524-1533 AND  
RP 1537-1544, SUBUNITS, AND INTERCHAIN DISULFIDE BOND.  
RC TISSUE=Serum;  
RX MEDLINE=93286045; PubMed=7685339;  
RA Oxvig C., Sand O., Kristensen T., Gleich G.J., Sottrup-Jensen L.;  
RT "Circulating human pregnancy-associated plasma protein-A is disulfide-  
RT bridged to the proform of eosinophil major basic protein."  
RT J. Biol. Chem. 268:12243-12246(1993).  
RN [4]  
RP PARTIAL SEQUENCE, CARBOHYDRATE-LINKAGE SITES, AND DISULFIDE BONDS.  
RX MEDLINE=22421368; PubMed=12421832;  
RA Overgaard M.T., Sorensen E.S., Stachowiak D., Boldt H.B.,  
RA Kristensen L., Sottrup-Jensen L., Oxvig C.,  
RT "Complex of pregnancy-associated plasma protein-A and the proform of  
RT eosinophil major basic protein. Disulfide structure and carbohydrate  
RT attachment sites."  
RT J. Biol. Chem. 278:2106-2117(2003).  
RN [5]

RP IDENTIFICATION, FUNCTION, SUBCELLULAR LOCATION, AND TISSUE  
 RP SPECIFICITY.  
 RC TISSUE=Fibroblast;  
 RX MEDLINE=99179030; PubMed=10077652;  
 RA Lawrence J.B., Oxvig C., Overgaard M.T., Sottrup-Jensen L.,  
 RA Gleich G.J., Hays L.G., Yates J.R. III, Conover C.A.;  
 RT "The insulin-like growth factor (IGF)-dependent IGF binding protein-4  
 RT protease secreted by human fibroblasts is pregnancy-associated plasma  
 RT protein-A";  
 RL Proc. Natl. Acad. Sci. U.S.A. 96:3149-3153(1999).  
 RN [6]  
 RP FUNCTION, SUBUNITS, AND ENZYME REGULATION.  
 RX MEDLINE=20469470; PubMed=10913121;  
 RA Overgaard M.T., Haaning J., Boldt H.B., Olsen I.M., Laursen L.S.,  
 RA Christiansen M., Gleich G.J., Sottrup-Jensen L., Conover C.A.,  
 RA Oxvig C.;  
 RT "Expression of recombinant human pregnancy-associated plasma protein-A  
 RT and identification of the proform of eosinophil major basic protein  
 RT as its physiological inhibitor";  
 RL J. Biol. Chem. 275:31128-31133(2000).  
 RN [7]  
 RP TISSUE SPECIFICITY.  
 RX MEDLINE=95057018; PubMed=7526035;  
 RA Bono M., Oxvig C., Kephart G.M., Wagner J.M., Kristensen T.,  
 RA Sottrup-Jensen L., Gleich G.J.;  
 RT "Localization of pregnancy-associated plasma protein-A and  
 RT ribonucleic acid and eosinophil granule major basic protein messenger  
 RT ribonucleic acid in placenta";  
 RL Lab. Invest. 71:560-566(1994).  
 RN [8]  
 RP TISSUE SPECIFICITY, AND DEVELOPMENTAL STAGE.  
 RX MEDLINE=99423540; PubMed=10491647;  
 RA Overgaard M.T., Oxvig C., Christiansen M., Lawrence J.B.,  
 RA Conover C.A., Gleich G.J., Sottrup-Jensen L., Haaning J.;  
 RT "Messenger ribonucleic acid levels of pregnancy-associated plasma  
 RT protein-A and the proform of eosinophil major basic protein:  
 RT expression in human reproductive and nonreproductive tissues";  
 RL Biol. Reprod. 61:1083-1089(1999).  
 RN [9]  
 RP DEVELOPMENTAL STAGE.  
 RX MEDLINE=95293954; PubMed=7539791;  
 RA Oxvig C., Haaning J., Kristensen L., Wagner J.M., Rubin I.,  
 RA Strigband T., Gleich G.J., Sottrup-Jensen L.;  
 RT "Identification of angiotensinogen and complement C3dg as novel  
 RT proteins binding the proform of eosinophil major basic protein in  
 RT human pregnancy serum and plasma";  
 RL J. Biol. Chem. 270:13645-13651(1995).  
 CC -1- FUNCTION: Metalloproteinase which specifically cleaves IGFBP-4 in  
 CC the presence of IGF, resulting in release of bound IGF.  
 CC -1- COFACTOR: Binds 1 zinc ion per subunit (BY similarity).  
 CC -1- ENZYME REGULATION: Inhibited by complexation with the proform  
 CC of PRG2.  
 CC -1- SUBUNIT: Homodimer; disulfide-linked. In pregnancy serum,  
 CC predominantly found as a disulfide-linked 2:2 heterotetramer with  
 CC the proform of PRG2.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- TISSUE SPECIFICITY: High levels in placenta and pregnancy serum.  
 CC in placenta, expressed in X cells in septa and anchoring villi,  
 CC and in syncytiotrophoblasts in the chorionic villi. Lower levels  
 CC are found in a variety of other tissues including kidney,  
 CC myometrium, endometrium, ovaries, breast, prostate, bone marrow,  
 CC colon, fibroblasts and osteoblasts.  
 CC -1- DEVELOPMENTAL STAGE: Present in serum and placenta during  
 CC pregnancy; levels increase throughout pregnancy.  
 CC -1- INDUCTION: By 8-bromadenosine-3',5'-phosphate.  
 CC -1- PTM: There appear to be no free sulfhydryl groups.  
 CC -1- SIMILARITY: Contains 5 Sushi (SCR) domains.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M46.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>  
 CC or send an email to [license@1sb-sib.ch](mailto:license@1sb-sib.ch)).  
 CC -----  
 CC EMBL; U28727; AAC50543.1; -;  
 CC EMBL; X68280; CAA48341.1; -;  
 CC PIR; S65464; S65464.  
 CC MEROPS; M46.001; -;  
 CC Genew; HGNC:8602; PAPPA.  
 CC MIM; 176385; -;  
 CC GO; GO:0008237; F:metallopeptidase activity; IDA.  
 CC GO; GO:0008270; F:zinc ion binding activity; NAS.  
 CC GO; GO:0007565; P:pregnancy; NAS.  
 CC InterPro; IPR006558; Lamg-like.  
 CC InterPro; IPR000800; Notch.  
 CC InterPro; IPR000436; Sushi\_SCR\_CCP.  
 CC InterPro; IPR006025; Zn\_Mtpeptidase.  
 CC Pfam; PF00084; sushi; 4.  
 CC SMART; SM00032; CCP; 4.  
 CC SMART; SM00560; lamgl; 1.  
 CC SMART; SM00004; NL; 3.  
 CC PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 CC KW Hydroxylase; Metalloproteinase; Metal-binding; Zinc; signal; Glycoprotein;  
 CC Zymogen; Repeat; Sushi.  
 CC FT SIGNAL 1 22  
 CC FT PROPEP 23 80  
 CC FT CHAIN 81 1627  
 CC FT DOMAIN 24 83  
 CC FT DOMAIN 272 583  
 CC FT DOMAIN 1215 1280  
 CC FT DOMAIN 1285 1342  
 CC FT DOMAIN 1346 1410  
 CC FT DOMAIN 1415 1471  
 CC FT DOMAIN 1478 1554  
 CC FT METAL 562 563  
 CC FT ACT SITE 563 563  
 CC FT METAL 566 566  
 CC FT METAL 566 566  
 CC FT DISULFID 144 235  
 CC FT DISULFID 327 622  
 CC FT DISULFID 332 657  
 CC FT DISULFID 414 428  
 CC FT DISULFID 424 440  
 CC FT DISULFID 427 473  
 CC FT DISULFID 461 461  
 CC FT DISULFID 474 485  
 CC FT DISULFID 583 600  
 CC FT DISULFID 587 612  
 CC FT DISULFID 710 710  
 CC FT DISULFID 713 881  
 CC FT DISULFID 732 732  
 CC FT DISULFID 753 835  
 CC FT DISULFID 775 781  
 CC FT DISULFID 947 975  
 CC FT DISULFID 960 971  
 CC FT DISULFID 983 990  
 CC FT DISULFID 999 1011  
 CC FT DISULFID 1036 1070  
 CC FT DISULFID 1051 1139  
 CC FT DISULFID 1192 1205  
 CC FT DISULFID 1210 1210  
 CC FT DISULFID 1215 1269  
 CC FT DISULFID 1227 1238  
 CC FT DISULFID 1242 1280  
 CC FT DISULFID 1285 1329  
 CC FT DISULFID 1300 1310  
 CC FT DISULFID 1314 1342  
 CC FT DISULFID 1346 1399  
 CC FT DISULFID 1362 1373  
 CC FT DISULFID 1377 1410  
 CC FT DISULFID 1415 1458  
 CC FT DISULFID 1428 1438  
 CC FT DISULFID 1442 1471  
 CC  
 CC INTERCHAIN (WITH C-51 OF PRG2 PROFORM).  
 CC OR 583-612.  
 CC OR 587-600.  
 CC  
 CC INTERCHAIN (WITH C-169 OF PRG2 PROFORM).  
 CC  
 CC INTERCHAIN.  
 CC



Query Match 39.7%; Score 3916.5; DB 1; Length 1627;  
 Best Local Similarity 45.8%; Pred. No. 9e-257;  
 Matches 718; Conservative 296; Mismatches 499; Indels 53; Gaps 25;

QY 249 REAETFMISQVGLP--IIYFSGRRRL--LIRPEVLAELPREAFYEAANVKEGGONPAII 305  
 DB 80 REARGATEBSPSPRALYFSGRGQLRVLRADL--ELPRDAFTIQWLRAEGGQSPAVI 137  
 QY 306 AGVFNDCSHTVSDGMALGIRSGKGRDAPRPFSLCTDPVKKATILISHRYQPQWT 365  
 DB 138 TGLYDKSYISRBKGVVGIHTTISDQNDKDRFFSLKTDRAOVYTTINARSTLPQOMV 197  
 QY 366 HVAATYDGRMALVYDGTQVASSLDQSGPLNSPFMACSRSLGDSSEBQHYRGLGT 425  
 DB 198 YLAATYDGMFKLVNQAQVATSGEYQGFISPLTQCKVLMIG--SALHNHVRGILIEH 255  
 QY 426 LVFWSTALPQSHFQSHSSGSEBEATDVLVTSFEPVNTWVPFRDEKYPRELV--LQG 483  
 DB 256 FSLMKVARTOREILSDMETHGAHTALPOLLOENWMDVVKHAMSFMKDGSSPKVFEFNAHG 315  
 QY 484 FEPEBELSLPQPLCGQTVCDNVELISQVNGVPLRGEKVIROYVNICDDDELNIVS 543  
 DB 316 FLAD---TSLPEPLCGQTLCDNTEVIASTYQLSSFPQPKVRYRVNVLXEDHKPFTVT 371  
 QY 544 EEOIIRLOHEALNENAFSRYNISWQSVHQNSTLRHVVLVNCEPSKIGNDHCDPECEHP 603  
 DB 372 REQVDFPHQDLAEFKQYINISWELDVLEVNSSLRRLILIANCISIKIGENDCEPCNHT 431  
 QY 604 LTGYDGDGR-LQGRCSYNNRDLGCHVECNMNLNDFDGDCCDQVADVAKTCEPDPSP 662  
 DB 432 LTGHGDCDCHLRPAFVKQHNVCMDCYERFNFDEGECDEPEITVNTQTCEPDDSP 491  
 QY 663 KRAYMSYKELKELQULSTHFTNLYFASVREDLAGAATPMDDAITHIGIYLSAYY 722  
 DB 492 HRAILDNENKNTIKLDGSTHNTLFFAKSSEBELAGVATPMDKALMLHIGIYLSNPFY 551  
 QY 723 GMEGHTDMHEVGHVGLYHVFKEVBERESCNPCKEIVPSMETGDLCADTAPFSEEL 782  
 DB 552 GMEGHTDMHEIGHSLGIVHVFGRGISEIQCSDPCMETESPFETGDLCDNTNAPRHS 611  
 QY 783 CREBEPTSDTCGTFTRPGAPFTNMYSTDNCTDNFTPNVARNHLYDLVYQWTSRK 842  
 DB 612 CGDPGPNDCGFEHSPFTPNYNNMSYADDCTDFTPNQARHNCYLDIVYQWQMSRK 671  
 QY 843 PTPIPIPMYIGQNKSLITHMLRPIGIVYVDRASGSLCAGACTDGFRRQVNHASRRV 902  
 DB 672 PAPVALAPQVLAGHTTDSVTLFEMPPIDGHFERELGASCHLCLEGRILVQVANSASPMP 731  
 QY 903 CDSGGYTPBEAVGPPIVDOPCEPSLOAMSPVHLHYMMNTVPCP--TEGSLLELFGHPV 961  
 DB 732 CSB9GHNSPREABEHPVDEOPCKSSVRTWSNANVNHVTPPACPEPQGCYLELEFYPL 791  
 QY 962 QADTLTLMVT--SPFMSSQVLPFTTEILLENKESVHLGPDTPCDIPYIKL--HYDKVS 1018  
 DB 792 VPESLTIWTVFVSTDMDSGAVNDIKLLAVSGKNISLGPQVFCDFVLTIRLMDVGEVY 851  
 QY 1019 GVKYTTDEKLEIDALITQSPHPLGSGCAPRVYQVLRDPPFASGLPVVYVTHSHRFTD 1078  
 DB 852 GIGIYTLDEHLEIDAAMLTSTADPRLCLQCKPLKXKVVRRPPLMDVASTIL--HINRFEVD 910  
 QY 1079 VEVTPGOMYQVYLABAGELGEASPLNHHGAPVGGDGVSRLEBECDDGLVSGDG 1138  
 DB 911 MDNLGSGVYQVWYTTISTGESSESPAVYTHHSGVGGDGIQXQDQEGCDMMKXINGDG 970  
 QY 1139 CSVKCELEBGFNCVGBPSLCMYEGDGICEPFEKTSIVDGIYTPKGYLDQWATRAYSS 1198  
 DB 971 CSIFCRQEVAFNCIDEBRSRCYFHDGQGVCEFEQTKSIKQGVYVTPGQFLDQWASNSVVS 1030  
 QY 1199 HEDKKKRPVSLVNGEP--HSLICTSYHNDLPHNRPLTGMFPVVAENETODDRSQPBGSL 1257  
 DB 1031 HQD-QQCPGAVIIGQAPASQVCRKVIIDLSEGISQAHMAYPCTTISYPSQ----- 1078

QY 1258 KKEDEWLVKVCENRPGEARAIFLFTTDLGVPEHQQPTVTLVTVDRGNSHSLGYGLS 1317  
 DB 1079 LAQTFPLRAVYFQSPVMAAVIYHLVLTDTGYDQGETISVQLDTRKQSHDLGLHVLVS 1138  
 QY 1318 QGNPLIINTHQNVLPHHTTSVLANFSSPRGISAVALRTSRLSLAPNSCISDEG 1377  
 DB 1139 CRNNPLIIPVNDHLSOPFYNSQAVRVSFSSPLVAISGVALRSPDNFPTLSSC-QRGET 1197  
 QY 1378 QNHQGSCHIRPCGKODSCPSLLDDHADVNTCSI---GPMKKAICITGORGALPASS 1433  
 DB 1198 YSPABGSCVFACEKTD-CPELAVENNS-LNCSSSDRYHG--AQTVCSTRTGYVLIR 1252  
 QY 1434 GQYIRPMQ--KEILLTSSGSHMDQNSCLPVDGVDPDPSLVNANFSCSGTFLKRCST 1491  
 DB 1253 DBELISQGTSPVYVTCETBKMKRYACEVDCSIDPHQVYVASFCEPQTTFGQCSF 1312  
 QY 1492 SCVPKALQGLSFWLTCLBGLMSLEPVYKLECDAPITLANLLPHCLQDNHVGTT 1551  
 DB 1313 QCRHPQLKGNNSILITCMEDGLMSFPEALCELMCLAPPPVPMADLQFARCRENHKVGFSF 1372  
 QY 1552 CKYCKRGYVVAASBGKVNKLKIQCLEGGIWEQSSCIPIVCEPPPVPEGVECTNG 1611  
 DB 1373 CKTKCKRGYHVPSSSR-KSKKRAPKTQCTODGSMQBGACVPTCDPPPKFHLGYCTNG 1431  
 QY 1612 FSLDQCVLNC-----NOEREKLPILCTKEGLTTOEFKLCENIQEGCPPPSBLNS-VGY 1665  
 DB 1432 FQNSBERICEDSDASQIGSNVHCRKQGTNTNGSHVQEQMGQC-SVPNELNSMLK 1490  
 QY 1666 KCEGIGIGAVCSPLCVTPSPDPVMLPENITADTLEHMMBPVKVQSVICTGRQMPDPV 1725  
 DB 1491 QCPDGYAIGSECATSCIDHNSSEIILPMYTVADIPIHMLPTVERVYCTAGLKMYPHA 1550  
 QY 1726 LVHCIOGCEPQADMCDDTNNRAYCHYDGDCCSSLTSSKXYIPFADDDLD-ECRCRD 1784  
 DB 1551 LHCVAGCEPMDGNYCDALNNRAPCNVDGDCCTSTVKTKVTPFPMSCDLQGDCACRD 1610  
 QY 1785 PKAEN 1790  
 DB 1611 PQAQEH 1616

RESULT 2  
 ID - CRL HUMAN STANDARD; PRT; 2039 AA.  
 AC P17927; Q16745; Q9UOV2;  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 01-NOV-1990 (Rel. 16, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Complement receptor type 1 precursor (C3b/C4b receptor) (CD35 antigen).  
 GN CRL OR C3BR.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.  
 NX NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=8903592; PubMed=2972794;  
 RA Klicstein L.B., Barlow T.J., Miletic V., Rabson L.D., Smith J.A.,  
 RA Fearon D.T.;  
 RT "Identification of distinct C3b and C4b recognition sites in the  
 RT human C3b/C4b receptor (CRL, CD35) by deletion mutagenesis.";  
 RL J. Exp. Med. 168:1699-1717(1988).  
 RN [2]  
 RP SEQUENCE FROM N.A. AND VARIANTS VAL-1615; ARG-1827 AND ASP-1850.  
 RX MEDLINE=94065175; PubMed=8245463;  
 RA Vuk D.P., Wong W.W.;  
 RT "Structure of the gene for the P allele of complement receptor type 1  
 RT and sequence of the coding region unique to the S allele.";  
 RL J. Immunol. 151:6214-6224(1993).  
 RN [3]  
 RP SEQUENCE OF 26-584 FROM N.A.  
 RX MEDLINE=89010527; PubMed=2971757;

RA Hourcade D., Miesner D.R., Atkinson J.P., Holers V.M.;  
 RT "Identification of an alternative polyadenylation site in the human  
 RT C3b/Cab receptor (complement receptor type 1) transcriptional unit and  
 RT prediction of a secreted form of complement receptor type 1.";  
 RL J. Exp. Med. 166:1255-1270(1988).  
 RN [4]  
 RP SEQUENCE OF 503-2039 FROM N.A.  
 RX MEDLINE=87168191; PubMed=2951479;  
 RA Klickestein L.B., Wong W.W., Smith J.A., Weis J.H., Wilson J.G.,  
 RA Fearon D.T.;  
 RT "Human C3b/Cab receptor (CRI). Demonstration of long homologous  
 RT repeating domains that are composed of the short consensus repeats  
 RT characteristics of C3/C4 binding proteins.";  
 RL J. Exp. Med. 165:1095-1112(1987).  
 RN [5]  
 RP SEQUENCE OF 761-783; 831-845 AND 1179-1195 FROM N.A.  
 RX MEDLINE=86067975; PubMed=2933745;  
 RA Wong W.W., Klickestein L.B., Smith J.A., Weis J.H., Fearon D.T.;  
 RT "Identification of a partial cDNA clone for the human receptor for  
 RT complement fragments C3b/C4b.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 82:7711-7715(1985).  
 RN [6]  
 RP VARIANTS ARG-1208; GLU-1590; GLY-1601; THR-1610; VAL-1615; ARG-1827  
 RP AND ASP-1850.  
 RX MEDLINE=21213564; PubMed=1131284;  
 RA Moulds J.M., Zimmerman P.A., Dounbo O.K., Kassambara L., Sagara I.,  
 RA Diallo D.A., Atkinson J.P., Krich-Goldberg M., Haubart R.B.,  
 RA Hourcade D.E., McNamara D.T., Birmingham D.J., Rowe J.A., Moulds J.J.,  
 RA Miller L.H.;  
 RT "Molecular identification of Knops blood group polymorphisms found in  
 RT long homologous region D of complement receptor 1.";  
 RL Blood 97:2879-2885(2001).  
 RN [7]  
 RP VARIANTS GLY-1601 AND THR-1610.  
 RX MEDLINE=21893860; PubMed=1189643;  
 RA Moulds J.M., Zimmerman P.A., Dounbo O.K., Diallo D.A., Atkinson J.P.,  
 RA Krich-Goldberg M., Hourcade D.E., Moulds J.J.;  
 RT "Expansion of the Knops blood group system and subdivision of SI(a).";  
 RL Transfusion 42:251-256(2002).  
 CC -1- FUNCTION: Mediates cellular binding of particles and immune  
 CC complexes that have activated complement.  
 CC -1- SUBUNIT: Monomer.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- TISSUE SPECIFICITY: Present on erythrocytes, leukocytes,  
 CC glomerular podocytes, and splenic follicular dendritic cells.  
 CC -1- POLYMORPHISM: CRI is responsible for the Knops (kn) blood group  
 CC system. It also carries the blood group antigens Mccoy (Mcc) and  
 CC Swain-Langley (Sl)/Villien (Vi).  
 CC -1- MISCELLANEOUS: Seven short consensus repeats (SCR) constitute a  
 CC long homologous repeat (LHR). The 2 N-terminal SRCS of LHR-A  
 CC contained a site determining C4 specificity, and the 2 N-terminal  
 CC SRCS of LHR-B and -C each had a site determining C3 specificity.  
 CC -1- MISCELLANEOUS: This is the sequence of the f allotype of CRI.  
 CC -1- SIMILARITY: BELONGS TO THE RECEPTORS OF COMPLEMENT ACTIVATION  
 CC (RCA) FAMILY.  
 CC -1- SIMILARITY: Contains 30 Sushi (SCR) domains.  
 CC -1- DATABASE: NAME=Blood group antigen mutation database;  
 CC NOTE=Knops (kn) blood group system;  
 CC WWW="http://www.bloc.aecom.yu.edu/bgmut/knops.htm".  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL: M11569; AAA52297.1; -  
 DR EMBL: L17418; AAB60694.1; -  
 DR EMBL: L17390; AAB60694.1; JOINED.  
 DR EMBL: L17399; AAB60694.1; JOINED.  
 DR EMBL: L17409; AAB60694.1; JOINED.

DR EMBL: L17419; AAB60694.1; JOINED.  
 DR EMBL: L17420; AAB60694.1; JOINED.  
 DR EMBL: L17421; AAB60694.1; JOINED.  
 DR EMBL: L17422; AAB60694.1; JOINED.  
 DR EMBL: L17423; AAB60694.1; JOINED.  
 DR EMBL: L17391; AAB60694.1; JOINED.  
 DR EMBL: L17392; AAB60694.1; JOINED.  
 DR EMBL: L17393; AAB60694.1; JOINED.  
 DR EMBL: L17394; AAB60694.1; JOINED.  
 DR EMBL: L17395; AAB60694.1; JOINED.  
 DR EMBL: L17396; AAB60694.1; JOINED.  
 DR EMBL: L17397; AAB60694.1; JOINED.  
 DR EMBL: L17398; AAB60694.1; JOINED.  
 DR EMBL: L17400; AAB60694.1; JOINED.  
 DR EMBL: L17401; AAB60694.1; JOINED.  
 DR EMBL: L17402; AAB60694.1; JOINED.  
 DR EMBL: L17403; AAB60694.1; JOINED.  
 DR EMBL: L17404; AAB60694.1; JOINED.  
 DR EMBL: L17405; AAB60694.1; JOINED.  
 DR EMBL: L17406; AAB60694.1; JOINED.  
 DR EMBL: L17407; AAB60694.1; JOINED.  
 DR EMBL: L17408; AAB60694.1; JOINED.  
 DR EMBL: L17410; AAB60694.1; JOINED.  
 DR EMBL: L17411; AAB60694.1; JOINED.  
 DR EMBL: L17412; AAB60694.1; JOINED.  
 DR EMBL: L17413; AAB60694.1; JOINED.  
 DR EMBL: L17414; AAB60694.1; JOINED.  
 DR EMBL: L17415; AAB60694.1; JOINED.  
 DR EMBL: L17416; AAB60694.1; JOINED.  
 DR EMBL: L17417; AAB60694.1; JOINED.  
 DR EMBL: X14362; CA32541.1; -  
 DR EMBL: M11617; AAA52298.1; -  
 DR EMBL: M11618; AAA52299.1; -  
 DR EMBL: X05309; CA28933.1; -  
 DR EMBL: X05309; CA28933.1; -  
 DR PIR: 173012; 173012.  
 DR PDB: 1GKX; 18-APR-02.  
 DR PDB: 1GKN; 18-APR-02.  
 DR PDB: 1GQP; 18-APR-02.  
 DR GeneW: HGNC:2334; CRI.  
 DR MIM: 120620; -  
 DR MIM: 607486; -  
 DR GO: GO:0005887; C:Integral to plasma membrane; TAS.  
 DR InterPro: IPR000436; Sushi\_SCR\_CCP.  
 DR Pfam: PF00084; sushi; 30.  
 DR SMART: SM00032; CCP; 30.  
 KW Complement pathway; Glycoprotein; Transmembrane; Repeat; Signal;  
 KW Receptor; Sushi; Blood group antigen; Polymorphism;  
 KW Pyroglutamate carboxylic acid; 3D-structure.  
 FT CHAIN 1 41  
 FT DOMAIN 42 2039  
 FT TRANSMEM 1972 1971  
 FT DOMAIN 1997 2039  
 FT MOD RES 42 42  
 FT DOMAIN 42 100  
 FT DOMAIN 103 162  
 FT DOMAIN 165 233  
 FT DOMAIN 237 294  
 FT DOMAIN 296 354  
 FT DOMAIN 357 417  
 FT DOMAIN 420 488  
 FT DOMAIN 492 550  
 FT DOMAIN 553 612  
 FT DOMAIN 615 683  
 FT DOMAIN 687 744  
 FT DOMAIN 746 804  
 FT DOMAIN 807 867  
 FT DOMAIN 870 938  
 FT DOMAIN 942 1000  
 FT DOMAIN 1003 1062  
 FT DOMAIN 1065 1133  
 FT DOMAIN 1137 1194  
 CC COMPLEMENT RECEPTOR TYPE 1.  
 CC EXTRACELLULAR (POTENTIAL).  
 CC POTENTIAL.  
 CC CYTOPLASMIC (POTENTIAL).  
 CC PYROGLUTAMATE CARBOXYLIC ACID (POTENTIAL).  
 CC SUSHI A1.  
 CC SUSHI A2.  
 CC SUSHI A3.  
 CC SUSHI A4.  
 CC SUSHI A5.  
 CC SUSHI A6.  
 CC SUSHI A7.  
 CC SUSHI B1.  
 CC SUSHI B2.  
 CC SUSHI B3.  
 CC SUSHI B4.  
 CC SUSHI B5.  
 CC SUSHI B6.  
 CC SUSHI B7.  
 CC SUSHI C1.  
 CC SUSHI C2.  
 CC SUSHI C3.  
 CC SUSHI C4.

```

FT DOMAIN 1196 1254 SUSHI C5.
FT DOMAIN 1257 1317 SUSHI C6.
FT DOMAIN 1320 1388 SUSHI C7.
FT DOMAIN 1395 1453 SUSHI D1.
FT DOMAIN 1456 1515 SUSHI D2.
FT DOMAIN 1518 1586 SUSHI D3.
FT DOMAIN 1590 1647 SUSHI D4.
FT DOMAIN 1649 1707 SUSHI D5.
FT DOMAIN 1710 1770 SUSHI D6.
FT DOMAIN 1773 1841 SUSHI D7.
FT DOMAIN 1847 1905 SUSHI E1.
FT DOMAIN 1908 1966 SUSHI E2.
FT DISULFID 43 86 BY SIMILARITY.
FT DISULFID 73 99 BY SIMILARITY.
FT DISULFID 104 145 BY SIMILARITY.
FT DISULFID 131 161 BY SIMILARITY.
FT DISULFID 166 215 BY SIMILARITY.
FT DISULFID 195 232 BY SIMILARITY.
FT DISULFID 238 280 BY SIMILARITY.
FT DISULFID 266 293 BY SIMILARITY.
FT DISULFID 297 340 BY SIMILARITY.
FT DISULFID 326 353 BY SIMILARITY.
FT DISULFID 358 400 BY SIMILARITY.
FT DISULFID 386 416 BY SIMILARITY.

```

Query Match 2.9%; Score 287.5; DB 1; Length 2039;

Best Local Similarity 19.1%; Pred. No. 5,9e-11;

Matches 349; Conservative 176; Mismatches 548; Indels 751; Gaps 105;

```

QY 454 LVLTASEPVNT-EWVP-----RDE-----KYRPLEVLOQEPPEPILSLQPL- 498
D 34 LALPVAAGCQNAPEMLFAPRTNLTDEFEPIGTLYNVECRPGVSGRPFSILCLKNSVMT 93
QY 499 -----CGQTYCDN-----VELISQYNGWPLRGKVI 525
D 94 GANDRCRKSCKRPPDPVNGVNHVIGIQQSGQIKYSCTKGYRLIGSSATCIISGDTVI 153
QY 526 RYQVNNICD--DEGLNPIVSEEQIRLQHEALNEAF--SRYNISMQLSVHGVHNSTLRHR 580
D 154 WDMETPICDRIPCGLPPTIT-----NGDFISTNRENFY-----GS 189
QY 581 VLVNCEPSKIGND-----HCDPECEHPLTGYDG--DCLQGRCYSMNRDGL 627
D 190 VVTVYRCNPGSGGRKVFELVGBPSIYCTSNDDQ--VGIMSGAPQCIIIPNKCTPPVNGEI 247
QY 628 CHVECNMM--LNDPDDGD--C-----DPOVADVAKTCEPDSPKRAY 666
D 248 LVSDNRSLFSLINEVEFRCPGFVWKGRPRRYKCOALNKMBELPSCSRVCQPPEDVLA- 306
QY 667 MSYKELKEALQLNSTHF--LNIYFASVREDLAGAATW-----PMDXAVT----- 710
D 307 -----ERQORDKDNFSFGQEVYVYSCPEGYDLGASMSRCTPGDMSPAAPICEVASC 359
QY 711 -----HLGGIVLSPAYYGMRGHTDTMHEVGHVLG--LYHVFKG-----VSERE 752
D 360 DFMGQLNGRVLPFVNLQAKVDVFCDEGQFKGSSASCVLGMSLNNSSVPCVEQI 419
QY 753 SCNDPCKEYTPS-METG-----DLCAD----- 773
D 420 FC--PSPBPVLPNGHRTKPLEVFPFGKAVNYTCDPHDRGTSFDLIGESTIRCTSDPOG 477
QY 774 ---TAPPKSEL--GREPS-----PTSDTCGTRPP--GAPFTYMS 808
D 478 GWMSSPAPRCGIIHGCAPDHFLPAKLTQTNASDFIGISLKECPREYVGRFS--- 533
QY 809 YTDNDCTDNF--TPNOVARMKCYLDLVYQOWTESRKPTPIPIPMV----- 852
D 534 ---ITCLDNLVWSPKVOVK-----RKSCTPPDPVAVGMHVITDIQVGRIN 578
QY 853 -----IGQYKSLTI-----HW--LPI-----SGVYVDASGLGACTEDGTF 891
D 579 YSCTTGHRLLIGHSSAECLISGNAHMSWKPEICORIPCGLPPTIANDPFI-----STNR 632

```

```

QY 892 QVYHTAS--SRVCDSSGYWTPEAVGPPDV-----DQCEPSLOAMS--PEVHLHYMM 942
D 633 ENHYGSVTVYRCNPGSGGRKVFELVGBPSIYCTSNDDQ-----VGIMSGAPQCIIIPNK 687
QY 943 TVCCPTEGSELLELPHQPVQADTLTLMTWSFFNESSQVLFDTLELLE--NKSVMHGL 999
D 688 CTPEPNVE-----NGILVSDNRSLFSLINEVEFRCPGFVWKGP 726
QY 1000 DTFCDIPLTIKLVQDKSVKYYTTPDERIEIDAALLTSQPHSLCSG--CRPVRYQLRD 1058
D 727 RVKQC-----ALNKBEPBELPSCSRVCQ----- 749
QY 1059 PPSASGLPVVVTSHRKFTDVE--VTEGOMYQYVLAEGEL--GEAS---PPLNHTIGA 1112
D 750 -----PPVLAHEQTQRKDNFSGQEVAFVS--GEQVYDLGASMSRCTPGDMSPA 800
QY 1113 PYCG-----DGK---VSRRLGE--CDDGDLVSGDGS----- 1140
D 801 PTECVKSCDDFMGQLNGRVLPFVNLQAKVDVFCDEGQFKGSSASCVLGMSLNM 860
QY 1141 ---KVGE-----LE-----EGFNCVGEPSL-C 1158
D 861 SSVPCEQIFCSPBPVLPNGHRTKPLEVFPFGKAVNYTCDPHDRGTSFDLIGESTIRC 920
QY 1159 YM-YEGDGI-----CE-----PFEKTSIVDCGIYT-----PKGY-- 1187
D 921 TSDPQNGVWSSPAPRCGIIHGCAPDHFLPAKLTQTNASDFIGISLKECPREYVGR 980
QY 1188 -----LDMATRAYSHED--KKYC--PVSIVTGBPSLI-----CTSYHPD 1225
D 981 PFSITCID--NLWSSPKDVCKRKSCKTPEPDVNGVNHVITDIQVSRINYCTTGH-R 1036
QY 1226 LPMNR-----PLTGMFC-----VASENETDDSEQEGSLKXEDV 1263
D 1037 LIGSSAECLISGNTAHMSTKPEICORIPCGLPPTIANGDFISTNRENFYGSV----- 1090
QY 1264 WLKVCENRPGEARAIF-----IFLTDDG-----LVGEHQOQTYT--LYL 1301
D 1091 -VTVYRCNLSGRKRVFELVGBPSIYCTSNDDQVIGISGAPQCIIIPNKCTPPVNGEILV 1149
QY 1302 TDVRSNHSI-----GTGLSCQ-----HNPILIIVTHHONVLF 1336
D 1150 SD-----NRSILFSLINEVEFRCPGFVWKGRPRRYKCOALNKMBELPSCSRVCQPP 1205
QY 1337 --HTTSVLNFSPPRGISAVLARTSIRGLSAPNSCISDEBQONOGOSCHIRPCKO- 1393
D 1206 GEHTPESHODNFS-----GOEVFYSC--EPGYDLGASLSH--CTPOG 1244
QY 1394 -----DSCPSEL--LDHADVYNCTSIGPLMKCAITCORGFALQASSGOYLRPM 1440
D 1245 DMSPEAPRCVAKSCDDPQLQLPHGRVLPPLNLQLG--AKVSFVCDDEGFRLKSSVSH- 1299
QY 1441 QKEILLTCSGHWDNVS--CLPVDGVPDPSLVNVA NFSCSG--TKFLKCSISCV- 1495
D 1300 ---CVLVGMSLNNSSVPCVEHIFCENP--PALNGRHGTGPSGDIPIYKKEISYTCDBHP 1355
QY 1496 ---PAKLGSLPULTLED--GLWSLREYVTKL-----ECAPPIILNANLLPRLCQ 1543
D 1356 RGMTFNLIGEST--IRCTSDPHGNGVWSSPAPRCLESVRACHKTPEOPFPASPPTIP--IN 1412
QY 1544 D-NHDVGTICKYCECKGYVVAESAEGKVRNKLKIOCLEGGIWE--QSGCIPVUCEP 1600
D 1413 DFEFPVGTSLANTCRGYF-----GKMSISCLENLVWSSVEDNKRKSSCEPPE 1462
QY 1601 VFEGM-----YECTNGFSL--DSQCVLNCQBEREKLPICTYEGMLTQ 1642
D 1463 PFMGMVHINTDQFGSTVNVSNCEGFRLLIGSPETTLCLVSGNNV-----TWDKK 1510
QY 1643 FKLCENLQCECPPEBELAS-----VEYKCEQYIG----- 1672
D 1511 APICEIIT--SCEPPTISNGDFPYSNNRTSFHNGTIVVYQCHTGPDBEQLELVGBRSIYC 1568
QY 1673 -----IGAVCSB-----LCVIPP--SDPVVLPENIT-----ADTLEHMEB- 1709

```

Db 1569 TSKDQGVWSSPPRCISTKCTAPAEVEMNIRVGNRSPSLREIIFRCQPGFVWVGS 1628  
Qy 1710 QSIYCTGRQRHPPVPLVHCQSCPE 1735  
Db 1629 HTVOCQTNGRW--GPKLPHCGRVCOP 1652

RESULT 3  
LEMS HUMAN  
ID LEM3 HUMAN STANDARD; PRT; 830 AA.  
AC P16109;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-AUG-1990 (Rel. 15, Last annotation update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE P-selectin precursor (granule membrane protein 140) (GMP-140) (PADGEM) (CD62P) (leukocyte-endothelial cell adhesion molecule 3) (LECAM3).  
GN SELP OR GMRP.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
NCBI\_TaxID=9606;  
CX (1)  
RX SEQUENCE FROM N.A.  
RX MEDLINE=89168432; PubMed=2466574;  
RA Johnson G.I., Cook R.G., McEver R.P.;  
RT "Cloning of GMP-140, a granule membrane protein of platelets and endothelium: sequence similarity to proteins involved in cell adhesion and inflammation."  
RT Cell 56:1033-1044(1989).  
RL (2)  
RX PALMITOYLATION.  
RX MEDLINE=93266599; PubMed=7684381;  
RA Fujimoto T., Stroud B., Whitley R.E., Prescott S.M., Muszbek L., Laposata M., McEver R.P.;  
RT "P-selectin is acylated with palmitic acid and stearic acid at cysteine 766 through a thioester linkage."  
RT J. Biol. Chem. 268:11394-11400(1993).  
RL (3)  
RX STRUCTURE BY NMR OF 160-199.  
RX MEDLINE=97057176; PubMed=8901515;  
RA Freedman S.J., Sanford D.G., Bachovchin W.W., Furie B.C., Baleja J.D., Furie B.;  
RT "Structure and function of the epidermal growth factor domain of P-selectin."  
RT Biochemistry 35:13733-13744(1996).  
RL (4)  
RX 3D-STRUCTURE MODELING OF 42-161.  
RX MEDLINE=94093388; PubMed=7505680;  
RA Bajorek J., Stenkamp R., Aruffo A.;  
RT "Knowledge-based model building of proteins: concepts and examples."  
RT Protein Sci. 2:1798-1810(1993).  
RL (5)  
RX VARIANTS ASN-331; ASP-603; VAL-640 AND PRO-756.  
RX MEDLINE=98334547; PubMed=9668170;  
RA Herrmann S.M., Ricard S., Nicaud V., Mallet C., Evans A., Rudavets J.B., Arveiler D., Luc G., Cambien F.;  
RT "The P-selectin gene is highly polymorphic: reduced frequency of the P7015 allele carriers in patients with myocardial infarction."  
RT Hum. Mol. Genet. 7:1277-1284(1998).  
RL (6)  
RX VARIANTS MET-209; LEU-301; ASN-331; VAL-365; PHE-500; ASP-603; VAL-640 AND PRO-756.  
RX MEDLINE=99318093; PubMed=10391209;  
RA Cargill M., Altshuler D., Ireland J., Sklar P., Ardlie K., Patil N., Shaw N., Lane C.R., Lim E.P., Kalyanaraman N., Nemes J., Ziaugra L., Friedland L., Rolfe A., Warrington J., Lipshutz R., Daley G.O., Lander E.S.;  
RT "Characterization of single-nucleotide polymorphisms in coding regions of human genes."  
RT Nat. Genet. 22:231-238(1999).  
RL (7)  
RX ERRATUM.

RA Cargill M., Altshuler D., Ireland J., Sklar P., Ardlie K., Patil N., Shaw N., Lane C.R., Lim E.P., Kalyanaraman N., Nemes J., Ziaugra L., Friedland L., Rolfe A., Warrington J., Lipshutz R., Daley G.O., Lander E.S.;  
RL Nat. Genet. 23:373-373(1999).  
CC -1- FUNCTION: CA(2+)-DEPENDENT RECEPTOR FOR MYELOID CELLS THAT BINDS TO CARBOHYDRATES ON NEUTROPHILS AND MONOCYTES. MEDIATES THE INTERACTION OF ACTIVATED ENDOTHELIAL CELLS OR PLATELETS WITH LEUKOCYTES. THE LIGAND RECOGNIZED IS SIALYL-LEWIS X.  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -1- TISSUE SPECIFICITY: STORED IN THE ALPHA-GRANULES OF PLATELETS AND WEIBEL-PALADE BODIES OF ENDOTHELIAL CELLS. UPON CELL ACTIVATION BY AGONISTS, P-SELECTIN IS TRANSPORTED RAPIDLY TO THE CELL SURFACE.  
CC -1- SIMILARITY: Contains 1 C-type lectin family domain.  
CC -1- SIMILARITY: Contains 1 EGF-like domain.  
CC -1- SIMILARITY: Contains 9 Sushi (SCR) domains.  
CC -1- DATABASE: NAME=PROV; NOTE=CD guide CD62P entry; WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd62p.htm".  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.ebi.ac.uk/announcements> or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
CC -----  
DR EMBL; M60234; AAA35910.1; JOINED.  
DR EMBL; M60217; AAA35910.1; JOINED.  
DR EMBL; M60218; AAA35910.1; JOINED.  
DR EMBL; M60219; AAA35910.1; JOINED.  
DR EMBL; M60222; AAA35910.1; JOINED.  
DR EMBL; M60223; AAA35910.1; JOINED.  
DR EMBL; M60224; AAA35910.1; JOINED.  
DR EMBL; M60225; AAA35910.1; JOINED.  
DR EMBL; M60226; AAA35910.1; JOINED.  
DR EMBL; M60227; AAA35910.1; JOINED.  
DR EMBL; M60228; AAA35910.1; JOINED.  
DR EMBL; M60229; AAA35910.1; JOINED.  
DR EMBL; M60231; AAA35910.1; JOINED.  
DR EMBL; M60232; AAA35910.1; JOINED.  
DR EMBL; M60233; AAA35910.1; JOINED.  
DR EMBL; M25322; AAA35911.1; JOINED.  
DR PDB; 1FSB; 01-APR-97.  
DR PDB; 1GIQ; 13-OCT-01.  
DR PDB; 1GIR; 13-OCT-01.  
DR PDB; 1G1S; 13-OCT-01.  
DR PDB; 1HES; 16-AUG-01.  
DR PDB; 1KJD; 03-APR-96.  
DR Genew; HGNC:10721; SELP.  
DR MIM; 173610; -.  
DR GO; GO:0005887; C:integral to plasma membrane; TAS.  
DR GO; GO:0005888; C:proteoglycan integral to plasma membrane; TAS.  
DR GO; GO:0005625; C:soluble fraction; TAS.  
DR GO; GO:0004895; F:cell adhesion receptor activity; TAS.  
DR InterPro; IPR006209; EGF-like.  
DR InterPro; IPR006210; IEGF.  
DR InterPro; IPR001304; Lectin\_C.  
DR InterPro; IPR002396; Selectin.  
DR InterPro; IPR000436; Sushi\_SCR\_CCP.  
DR Pfam; PF00008; lectin\_1.  
DR Pfam; PF00059; lectin\_C\_1.  
DR Pfam; PF00084; sushi\_9.  
DR PRINTS; PR00343; SELECTIN.  
DR SMART; SM00032; CCP; 9.  
DR SMART; SM00034; CLECT; 1.  
DR SMART; SM00181; EGF; 1.  
DR PROSITE; PS00022; EGF\_1; 1.  
DR PROSITE; PS01186; EGF\_2; 1.  
DR PROSITE; PS00615; C\_TYPE\_LLECTIN\_1; 1.  
DR PROSITE; PS00615; C\_TYPE\_LLECTIN\_2; 1.  
DR PROSITE; PS00615; C\_TYPE\_LLECTIN\_2; 1.

KW Cell adhesion; Transmembrane; Glycoprotein; EGF-like domain; Lectin;  
 KW Selectin; Signal; Sushi; Repeat; Lipoprotein; Palmitate; Polymorphism;  
 KW 3D-structure.

FT	SIGNAL	1	41
FT	CHAIN	42	830
FT	DOMAIN	42	771
FT	TRANSMEM	772	795
FT	DOMAIN	796	830
FT	DOMAIN	58	158
FT	DOMAIN	159	195
FT	DOMAIN	199	258
FT	DOMAIN	261	320
FT	DOMAIN	323	382
FT	DOMAIN	385	444
FT	DOMAIN	447	506
FT	DOMAIN	509	568
FT	DOMAIN	571	630
FT	DOMAIN	641	700
FT	DOMAIN	703	762
FT	DISULFID	60	158
FT	DISULFID	131	150
FT	DISULFID	163	174
FT	DISULFID	168	183
FT	DISULFID	185	194
FT	DISULFID	200	244
FT	DISULFID	230	257
FT	DISULFID	262	306
FT	DISULFID	292	319
FT	DISULFID	324	368
FT	DISULFID	354	381
FT	DISULFID	386	430
FT	DISULFID	416	443
FT	DISULFID	448	492
FT	DISULFID	478	505
FT	DISULFID	510	554
FT	DISULFID	540	567
FT	DISULFID	572	616
FT	DISULFID	602	629
FT	DISULFID	642	686
FT	DISULFID	672	699
FT	DISULFID	704	748
FT	DISULFID	734	761
FT	CARBOHYD	54	58
FT	CARBOHYD	98	98
FT	CARBOHYD	180	180
FT	CARBOHYD	212	212
FT	CARBOHYD	219	219
FT	CARBOHYD	411	411
FT	CARBOHYD	460	460
FT	CARBOHYD	518	518
FT	CARBOHYD	665	665
FT	CARBOHYD	716	716
FT	CARBOHYD	723	723
FT	CARBOHYD	741	741
FT	LIPID	807	807
FT	SITE	818	821
FT	VARIANT	209	209
FT	VARIANT	301	301
FT	VARIANT	331	331
FT	VARIANT	365	365

Query Match 2.5%; Score 249; DB 1; Length 830;  
 Best Local Similarity 22.1%; Pred. No. 6.4e-09;  
 Matches 102; Conservative 59; Mismatches 160; Indels 140; Gaps 25;

QY 1382 GQSCIH-RPGKQDSCSLLDHADVNCISIGBGLMKALTCORGALDASSGQYIRPM 1440  
 DB 191 GPCEYVREGGELELPQHVILMNCSHPLGNFSFN---SQCSFHCTDGYQVNG-----PS 240

QY 1441 QKEILLTSCSGHW-DONVCLPVDCGVDPDSLVNANFSCSEBCKFLKR---CSISC--- 1493  
 DB 241 KLECL---ASGITNNKPPQCLAAQC---PLKIPERGNMILMSAKAFHQSSCSFSCBEG 295  
 QY 1494 ---VPAKIQGLSPWLTCLDEGLMSLPEYCK-LBC---DAP-----PIILANLL 1537  
 DB 296 FALVGEVQV-----CTASGWTAPAPVCAVQCCHLEAPSEGTMDCVHPILATA--- 345  
 QY 1538 LPHCLDNDHNVGICIKYCKECPGYVVASBAGKRNLLKLCLEGGIWEQ--GSCI PVVC 1595  
 DB 346 -----YSSCKFECOPG-----RVRG-LDMRCITDSGHSAPLPFCATISC 386  
 QY 1596 EPPVPVEGVYECT---NGFSLDSQCLVNCQE---REKLPIVCTKRGMLTOEFKCENTL 1649  
 DB 387 EPLSPVHSGMDCSPBLRAFQYDTNCSFRCABGFMRLGADIVACDNLGGWTAPAPVQAL 446  
 QY 1650 QGECPPPESELSVRYKCEQVY--IGAVCS-----P 1679  
 DB 447 QCCDLVFP---NEARVNCSPHFGAFRYQSVCSFTCNELGLLVGASVLOCLATGNMNSVVP 503  
 QY 1680 LCVIPSPDPMLEPENTADTLHEHMEPVKVS-----IVCTGRQ 1719  
 DB 504 ECAIPCTPULSPFQNGTMT---VPLGSSSYKSTCQFICDEGYLSGPERLCTRSR 559  
 QY 1720 WHDPVLVACIOGCEPFQAD-GWCDTINNRR-----AYCHY 1753  
 DB 560 WIDSPMCAIKCELPFAPEQSSLDSDTRGERFNVSTCHF 600

RESULT 4  
 CR2\_HUMAN STANDARD; PRT: 1033 AA.  
 ID P20523; O13866; Q14212;  
 AC 01-FEB-1991 (Rel. 17, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Complement receptor type 2 precursor (Cr2) (Complement C3d receptor)  
 DE (Epstein-Barr virus receptor) (EBV receptor) (CD21 antigen).  
 GN CR2 OR C3DR.  
 OS Homo sapiens (human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP MEDLINE=8913277; PubMed=2563370;  
 RA Fujisaku A., Harley J.B., Frank M.B., Gruner B.A., Frazier B.,  
 RT "Genomic organization and polymorphisms of the human C3d/Epstein-Barr  
 RT virus receptor.";  
 RL J. Biol. Chem. 264:2118-2125(1989).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RP TISSUE=B-cell;  
 RC MEDLINE=88171282; PubMed=2832506;  
 RA Weiss J.J., Tothaker L.E., Smith J.A., Weiss J.H., Fearon D.T.;  
 RT "Structure of the human B lymphocyte receptor for C3d and the  
 RT Epstein-Barr virus and relatedness to other members of the family  
 RT C3/C4 binding proteins.";  
 RL J. Exp. Med. 167:1047-1066(1988).  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOPFORM C).  
 RP MEDLINE=88097454; PubMed=2827171;  
 RA Moore M., Cooper N., Tack B., Nemerow G.;  
 RT "Molecular cloning of the cDNA encoding the Epstein-Barr virus C3d  
 RT receptor (complement receptor type 2) of human B lymphocytes.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 84:9194-9198(1987).  
 RN [4]  
 RP SEQUENCE FROM N.A. (ISOPFORMS A; C AND D).  
 RP MEDLINE=99165388; PubMed=10068037;  
 RA Baril M., Balbo M., Frade R.;  
 RT "Evidence for a new transcript of the Epstein-Barr virus C3d receptor  
 RT (CR2, CD21) which is due to alternative exon usage.";



```

FT  VARSPIC 659 659 /Frid=VSP_001209.
FT  K -> KGCOPBPGHHRHGTGNTVPFVSGATVDTGDDPX
FT  LVNGNSICHMPSGNSMSPSPARCE (in isoform C
FT  and isoform D).
FT  VARSPIC 716 723 /Frid=VSP_001210.
FT  Missing (in isoform D).
FT  CONFLICT 457 457 /Frid=VSP_001211.
FT  CONFLICT 646 646 MISSING (IN REF. 2).
FT  CONFLICT 667 667 A -> R (IN REF. 3).
FT  CONFLICT 759 787 Q -> D (IN REF. 5).
FT  KLCGRSDKSHGWSGSPQCLRRPVTNR -> NCSAEVIL
FT  KAMLERAPQCLRRSL (IN REF. 3).

Query Match 2.5%; Score 249; DB 1; Length 1033;
Best Local Similarity 20.3%; Pred. No. 8.8e-09;
Matches 227; Conservative 115; Mismatches 400; Indels 374; Gaps 62;

776 PTKSELRE-----PEPTSDTCGTPRPPGAPFTNYSYTDNCTDNTFNQVARMHCYLD 831
78 PARKCEYFNKYSCEPIVPG-CYKIRGSTPRRGDSVT-PAKTNFSMNGKNSVWCQAN 135
832 LTVQQTESKPT-----PIPIP--PMVIGQTKSLTIHMLPPISGVYDRAAGSLGAC 884
136 ---NMGPRLPTCVSVFPLECPALPMIHNGHSTENVGSIAPLSVTSGCEG----- 186
885 TDEGTFOYVHTASSRVCDSGWTPEAVGPRDVDPCE-----PSLOAMSEVHLX 939
187 -----YLVGKTIINCLSSGKMS--AVPPCEARCKSLGRFNGKVEPIIRVG 235
940 MNMTVPCTEGCSLELLEFQHP-----VQADTLTLNV-----TSFFMSSQVLFETILE 969
236 VTANFPC-DEGYRL-----QGPSSRCVIAOGVAMTQMPCEIFCSPPIILNGRIHN 290
990 NKESVHLGLDPT-CDI-----PLTIKLVHDKVSGVKYTFD--ERLEIDA 1033
291 SLANVSYSIVTYCTDDEPREGVNFILIGESTLACTYDSQKTG---TWGSPARCELST 346
1034 ALLTQSHPSLSCGCRVRVQVLRDPPFASGLPVVTHSHRKFTDVEVTGQVQY----- 1089
347 SAVQC-PHP-----QILR-----GRWVGSKDRYTYNDT-VIPACMFGLKG 387
1090 --OVLAEAGELGASPLNHIHAPYCGDKVSERLGECDDDLVSGDCSVCELE 1147
388 SKQIRCAOGTWESAP-----VAACEATGRQLLTCKRQHQ-----VCCKE- 409
1148 GFNCVGEPSLCMYBGDICEPERKTSIVDCGIYTERGYLDQWATRAYSHEDKXC-P 1206
410 ---QAPRNILNGCKEDRHMYRFPDGTSL-----KYSQNP 441
1207 VSLVGTGPHSLICTS---YHPDLNHRPLTGMFPCVASENETODDRSEORPBGSLKXDEV 1263
442 GYVALVG-ESIQCTSEGVMTPPVQCK-----VAACEATGRQLLTCKRQHQ----- 485
1264 WLKCFRFP-----GEARAI-FIPLTDLGVGEHQOFTVLYLTDVGSNHSLSGTYG 1315
486 -----FVRDVNSGCEBGYKLSGSVYQECGTI-----WFEHILCKE----- 524
1316 LSCQHNPLIINVTHQNLV---FNHTTSVLNFS-SPRVGISAVALRTSSRIGLSAPNCI 1372
525 ITCPPRPVINYGAHTGSLDEFPYGTIVTYTCNGBRGVE-----FSLIGEST-IRCT 577
1373 SDEGQNHQOQSCIHRCQKQDSCPSLL---DHAVVNCSTI-----GGLMKCAIT-- 1422
578 SNOQERTWS-----GPAPLCKSLLVAVQCSHVIAIANGYKLSGKEAPFYNDVTFK 629
1423 CORGFALQASSGOYIRPMOKELILTCCSGH-WDQNVG-CLFVPC-----GVDPDSLV 1472
630 CYSGFTLKSS-----QIRCADNTWDEIIVCEKERTQHRQSLQELPSSRV 678
1473 NYANFSGSBGTFKLRCSISCVPAKIQGSLFWLTCLADGLMSLPEVYCK-LBCDAPPI 1531
679 ELVNTSCODGYQLTGHAYQWQOD-----AENGIMFKPIKICKVIHCHPPPIV 725

```

```

QY 1532 LMANLLLP--CLQDNHDVGTICKYCECKPGYVAESAEGVRNKKLKICQLEG-----GIW 1585
DB 726 VNGK-----HTGMMAENPLVNGEVSYECDQGFYLGK-----KLCGRSDSKHGSW 772
QY 1586 EOGS--CI---PVYCEPPPPVPEG-----MYECTNGFSLDSCVLMNCQ 1624
DB 773 SGSPQCLSRPPVTRCPNPEVKGKYLKNTKSHAYSHNDIYVVDNCNGFIANGSRVIRCHT 832
QY 1625 EREKTLPL--CTYEGIMTQEFKLCENLQCECPPEPSELN-----SYEYKC 1667
DB 833 DNTWVGVPCTICKAPF-----GCPPPPTPNMHTGNIARSPGMSILVSC 880
QY 1668 EOGYGIQAVCSPLCV-----IPSPDPVMLPENITADTLEHMEPVRKYQ---STV--- 1713
DB 881 DQGITLVGEALILCTHGTGSPAPHCYKCVNCSPPADMSTIQNGLEBRKRYQYGAVTTLE 940
QY 1714 -----CTGRQRQHPDPVYLWHCIQSCPEP 1735
DB 941 CEDGYMLGSSPQSQCSQSDHQMNPP-PLAVCRSRLAP 975

RESULT 5
LEW2 MOUSE
ID LEW2 MOUSE STANDARD; PRT; 612 AA.
AC 000650:
DT 01-ARR-1993 (Rel. 25, Created)
DT 01-ARR-1993 (Rel. 25, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE E-selectin precursor (Endothelial leukocyte adhesion molecule 1)
DE (ELAM-1) (Leukocyte-endothelial cell adhesion molecule 2) (LECAM2)
DE (CD62E).
DE SELE OR ELAM-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92283265; PubMed=1375914;
RA Becker-Andre M., van Huijsduijnen R.H., Losberger C., Whelan J.,
RA Delamarier J.F.;
RT "Mxine endothelial leukocyte-adhesion molecule 1 is a close
RT structural and functional homologue of the human protein."
RL Eur. J. Biochem. 206:401-411(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92340571; PubMed=1378846;
RA Weller A., Isenmann S., Vestweber D.;
RT "Cloning of the mouse endothelial selectins. Expression of both E-
RT and P-selectin is inducible by tumor necrosis factor alpha."
RL J. Biol. Chem. 267:15176-15183(1992).
CC - FUNCTION: EXPRESSED ON CYTOKINE INDUCED ENDOTHELIAL CELLS AND
CC MEDIATES THEIR BINDING TO LEUKOCYTES. THE LIGAND RECOGNIZED BY
CC ELAM-1 IS STAYL-Lewis X (ALPHA(1)-3) FUCCOSYLATED DERIVATIVES OF
CC POLYACTOSAMINE THAT ARE FOUND AT THE NONREDUCING TERMINI OF
CC GLYCOTIDES.
CC - SUBCELLULAR LOCATION: Type I membrane protein.
CC - SIMILARITY: BELONGS TO THE SELECTIN/LECAM FAMILY.
CC - SIMILARITY: Contains 1 C-type lectin family domain.
CC - SIMILARITY: Contains 1 BGF-like domain.
CC - SIMILARITY: Contains 6 Sushi (SCR) domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL; M80778; AAA37547.1;
CC EMBL; M87862; AAA37577.1; ALT INIT.
DR PIR; S23174; B42755.

```

```

DR HSPB; P16581; IKJA.
DR MED; MGI:98278; Sele.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR003304; Lectin C.
DR InterPro; IPR002396; Selectin.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00008; EGF_1.
DR Pfam; PF00059; lectin_C_1.
DR Pfam; PF00084; sushi_6.
DR PRINTS; PR00343; SELECTIN.
DR SMART; SM00032; CCP_6.
DR SMART; SM00034; CLECT_1.
DR SMART; SM00181; EGF_1.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS00615; C-TYPE LECTIN 1; 1.
DR PROSITE; PS50041; C-TYPE LECTIN 2; 1.
DR Cell adhesion; Transmembrane; Glycoprotein; EGF-like domain; Lectin;
KM Selectin; Signal; Sushi; Repeat.
FT SIGNAL 1 21 BY SIMILARITY.
FT CHAIN 22 612 E-SELECTIN.
FT DOMAIN 22 557 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 558 579 POTENTIAL.
FT DOMAIN 580 612 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 38 138 C-TYPE LECTIN (SHORT FORM).
FT DOMAIN 139 175 EGF-LIKE.
FT DOMAIN 179 239 SUSHI 1.
FT DOMAIN 242 301 SUSHI 2.
FT DOMAIN 304 364 SUSHI 3.
FT DOMAIN 367 427 SUSHI 4.
FT DOMAIN 430 490 SUSHI 5.
FT DOMAIN 493 549 SUSHI 6.
FT DISULFID 40 138 BY SIMILARITY.
FT DISULFID 111 130 BY SIMILARITY.
FT DISULFID 143 154 BY SIMILARITY.
FT DISULFID 148 163 BY SIMILARITY.
FT DISULFID 165 174 BY SIMILARITY.
FT DISULFID 180 225 BY SIMILARITY.
FT DISULFID 210 238 BY SIMILARITY.
FT DISULFID 243 287 BY SIMILARITY.
FT DISULFID 273 300 BY SIMILARITY.
FT DISULFID 305 350 BY SIMILARITY.
FT DISULFID 336 363 BY SIMILARITY.
FT DISULFID 368 413 BY SIMILARITY.
FT DISULFID 399 426 BY SIMILARITY.
FT DISULFID 431 476 BY SIMILARITY.
FT DISULFID 462 489 BY SIMILARITY.
FT DISULFID 494 535 BY SIMILARITY.
FT DISULFID 521 548 BY SIMILARITY.
FT CARBOHYD 25 25 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 145 145 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 192 192 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 203 203 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 266 266 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 333 333 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 391 391 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 504 504 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 528 528 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 612 AA; 66749 MW; 86F05713F0EC2C3D CRC64;

Query Match 2.5; Score 247; DB 1; Length 612;
Best Local Similarity 21.8; Pred. No. 5.6e-09;
Matches 121; Conservative 60; Mismatches 225; Indels 148; Gaps 27;
1233 TGWPCVASENETODRSEQPEG-----SLKKEDE 1262
20 TAWYNNASSELMYDEASAYCQRDYTHVAIONKEINIYNSLNKHSPIYWIGIRKYN 79
1263 VMLKVCNRPGEARAFIFLTTDGLVPE-----HQPTVLYLTDVAGS----- 1307
80 VMIVWGCKPLTEA-----QWAPGEPNKKORNEDEVEIYIORTKDSGMWDERCNK 132

```

```

QY 1308 -----NHSCTYGV-----LSCQNPILINVTNHNQVLF-----HHTSVLNLN 1344
DB 133 KUALCTYASCNTNNSGSHGECIETTINSYTCCKHGLGNCEDAVTCRQEHDPYSLN 192
QY 1345 FSSPRVGSVALARTSSRIGLS--APSN-----CISEDEGONHQSCIRHPCGKQDS 1395
DB 193 CSHP---FGPFSYNSGSCFCGCKRGYLPSSMETTYRCTS-----SGEMGAPAPACHIVE 242
QY 1396 CPSLLDHDADVNTSGIPGL-----MKCAITCQGRFALQASGQYIRPMQKEILLTSSG 1451
DB 243 CEALTHPAHGIKRCSS--NPSPYMNNTCTPDCEVG-----RRVAQNLQCTSSG 291
QY 1452 HMD-QNVSCLPVDC-GVPDPSLVYANFSGSEGT-----KFLKRCISGVPPAKQLGSPW 1505
DB 292 IMDNETPSCAVYTCDAIPQ---QNGFVSCSHSTAGLAKSSCNFTCEGSPFLQGPQ- 347
QY 1506 LTYLEDLWSLPEVYCK-LBCDAPPIILNALLPHCLQDNHDVGTICKEYCKGYVAE 1564
DB 348 VEGSAQGWTPQIPVCAVQCEALAPQGNMKCLPSASBPFGQSSCFSCGEGFELKG 407
QY 1565 SAGSKVKNKLIKQCLEGGIWE--QSSCIPVCEPPPPVEGMYEC-----TNGFSLDSOC 1618
DB 408 SR-----RLQCGPRGEMDSKKPTCSAVKCDVDPRPQNGWECNAHTGTEPTYSSC 458
QY 1619 VLNCNOE---REKPLICTKGLWTOBFKLCENLQGECP--PPSELNSVEYKCEQGYGI 1673
DB 459 AFQCNBEFSLHGAQLCTCSGKKTQDEVPSQVQV--CPSLDVPGKKN--MSCSGTAVF 513
QY 1674 GAVCSPLCVIIPSD 1687
DB 514 GTVCEFTC---PDD 524

RESULT 6
CPAH MOUSE STANDARD; PRT; 1234 AA.
ID P06909;
AC 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Complement factor H precursor (Protein beta-1-H).
GN HFI OR CFH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=6623353; PubMed=2940596;
RA Kristensen T., Tack B.P.;
RT "Murine protein H is comprised of 20 repeating units, 61 amino acids
RT in length.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:3963-3967(1986).
RP [2]
RP SEQUENCE OF 1-19 FROM N.A.
RX STRAIN=BALB/c;
RX MEDLINE=90148935; PubMed=2533512;
RA Munoz-Carover P., Tack B.F., Vix D.P.;
RT "Analysis of complement factor H mRNA expression: dexamethasone and
RT IFN-gamma increase the level of H in L cells.";
RL Biochemistry 28:9891-9897(1989).
RN [3]
RP SEQUENCE OF 1-18 FROM N.A.
RX MEDLINE=9011033; PubMed=2136885;
RA Natsume-Sakai S., Nonaka M., Harada Y.N., Shreffler D.C.,
RA Moriwaki K.;
RT "Demonstration of an unusual allelic variation of mouse factor H by
RT the complete cDNA sequence of the H.2 allotype.";
RL J. Immunol. 144:358-362(1990).
CC -1- FUNCTION: Factor H functions as a cofactor in the inactivation of
CC C3b by factor I and also increases the rate of disassociation of the
CC C3bb complex (C3 convertase) and the (C3b)NB complex (C5
CC convertase) in the alternative complement pathway (By similarity).

```



CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- TISSUE SPECIFICITY: Synthesized by liver and secreted in plasma.  
 CC -1- POLYMORPHISM: TWO CODOMINANT ALLELES OF FACTOR H ARE PRESENT IN MICE.  
 CC -1- SIMILARITY: Contains 20 Sushi (SCR) domains.  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.ebi.ac.uk/announcements> or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 -----  
 CC EMBL: M12660; AAA37759.1; -;  
 DR EMBL: J02891; AAA37795.1; -;  
 DR EMBL: M31979; AAA37762.1; -;  
 DR PIR: A26154; NEMSH.  
 DR HSSP: P08603; 1HFI.  
 DR MGD: MGI:88385; Cfh.  
 DR InterPro: IPR00436; Sushi\_SCR\_CCP.  
 DR Pfam: PF00084; sushi; 20.  
 DR SMART: SM0032; CCP; 20.  
 KM Complement alternate pathway; Plasma; Glycoprotein; Repeat; Sushi; Signal.  
 FT SIGNAL 1 18 BY SIMILARITY.  
 FT CHAIN 19 1234 COMPLEMENT FACTOR H.  
 FT DOMAIN 20 81 SUSHI 1.  
 FT DOMAIN 84 142 SUSHI 2.  
 FT DOMAIN 145 206 SUSHI 3.  
 FT DOMAIN 209 263 SUSHI 4.  
 FT DOMAIN 266 321 SUSHI 5.  
 FT DOMAIN 324 386 SUSHI 6.  
 FT DOMAIN 388 443 SUSHI 7.  
 FT DOMAIN 447 506 SUSHI 8.  
 FT DOMAIN 508 565 SUSHI 9.  
 FT DOMAIN 568 623 SUSHI 10.  
 FT DOMAIN 628 684 SUSHI 11.  
 FT DOMAIN 689 744 SUSHI 12.  
 FT DOMAIN 751 803 SUSHI 13.  
 FT DOMAIN 807 862 SUSHI 14.  
 FT DOMAIN 866 932 SUSHI 15.  
 FT DOMAIN 935 990 SUSHI 16.  
 FT DOMAIN 993 1049 SUSHI 17.  
 FT DOMAIN 1052 1108 SUSHI 18.  
 FT DOMAIN 1113 1169 SUSHI 19.  
 FT DOMAIN 1171 1234 SUSHI 20.  
 FT DISULFID 21 66 BY SIMILARITY.  
 FT DISULFID 52 80 BY SIMILARITY.  
 FT DISULFID 85 129 BY SIMILARITY.  
 FT DISULFID 114 141 BY SIMILARITY.  
 FT DISULFID 146 192 BY SIMILARITY.  
 FT DISULFID 178 205 BY SIMILARITY.  
 FT DISULFID 210 251 BY SIMILARITY.  
 FT DISULFID 237 262 BY SIMILARITY.  
 FT DISULFID 267 309 BY SIMILARITY.  
 FT DISULFID 294 320 BY SIMILARITY.  
 FT DISULFID 325 374 BY SIMILARITY.  
 FT DISULFID 357 385 BY SIMILARITY.  
 FT DISULFID 389 431 BY SIMILARITY.  
 FT DISULFID 416 442 BY SIMILARITY.  
 FT DISULFID 448 494 BY SIMILARITY.  
 FT DISULFID 477 505 BY SIMILARITY.  
 FT DISULFID 509 553 BY SIMILARITY.  
 FT DISULFID 536 564 BY SIMILARITY.  
 FT DISULFID 569 610 BY SIMILARITY.  
 FT DISULFID 597 622 BY SIMILARITY.  
 FT DISULFID 629 672 BY SIMILARITY.  
 FT DISULFID 658 683 BY SIMILARITY.  
 FT DISULFID 690 732 BY SIMILARITY.  
 FT DISULFID 718 743 BY SIMILARITY.  
 FT DISULFID 752 791 BY SIMILARITY.  
 FT DISULFID 780 802 BY SIMILARITY.

FT DISULFID 808 850 BY SIMILARITY.  
 FT DISULFID 836 861 BY SIMILARITY.  
 FT DISULFID 867 920 BY SIMILARITY.  
 FT DISULFID 906 931 BY SIMILARITY.  
 FT DISULFID 936 978 BY SIMILARITY.  
 FT DISULFID 964 989 BY SIMILARITY.  
 FT DISULFID 994 1037 BY SIMILARITY.  
 FT DISULFID 1023 1048 BY SIMILARITY.  
 FT DISULFID 1053 1096 BY SIMILARITY.  
 FT DISULFID 1082 1107 BY SIMILARITY.  
 FT DISULFID 1114 1157 BY SIMILARITY.  
 FT DISULFID 1143 1168 BY SIMILARITY.  
 FT DISULFID 1172 1223 BY SIMILARITY.  
 FT DISULFID 1206 1233 BY SIMILARITY.  
 FT CARBOHYD 676 676 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 721 721 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 773 773 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 801 801 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1030 1030 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1061 1061 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1225 1225 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 1234 AA; 139082 MW; C5AC02P341B957F7 CRC64;  
 -----  
 Query Match 2.4%; Score 238.5; DB 1; Length 1234;  
 Best Local Similarity 17.8%; Pred. No. 5.9e-08;  
 Matches 144; Conservative 83; Mismatches 239; Indels 343; Gaps 35;  
 -----  
 QY 1181 IYTPKGIIDOMATRAVSHEDKKCPVSLV--TGE-PHSLICTSYHEDLPNHRPLTGWFP 1237  
 D 453 IHDNGFLSSSSSYALNRETSYCKQGYVNTGHSITCQ-----NGMSP 501  
 QY 1238 ---CVASENETQDDBRSP--EGSLKKEDEWLKVCNRRGEARAF--IFLTDTGLVPG 1290  
 D 502 QPSCIKS-----CMPVFNISITNTRTWFLDNKLDECLVGEENEYKHTKSTI-- 551  
 QY 1291 EHQQPTVTLVLTDRGSHSLAGTYGLSCQ---HNPLIYNTHQNVLFPHHTSVLLNF- 1345  
 D 552 -----TCTY-----GMSDTPSCYERECVPTLDRLKLVSPRKEK-----YRVGDLLEFS 596  
 QY 1346 --SSPRGISAVNL-----RTSSRIGLSPNSCISEDEQNHQOGSCIRPGCKQ 1393  
 D 597 CHSGHRVGPDSVOCYHFGWSPGPPTCKGVASCAPLEIINGELNGAKVEYSHGEVVKY 656  
 QY 1394 DSCPSLLDHDADVNCI-----SIGPGLMKCAI----- 1421  
 D 657 DCKPRFLKGPNNKIQCVDGNWTLTPVCIEBERTCGDIPELHSGAKSVPRYHHGDSVER 716  
 QY 1422 TCQGFALQASSGQYIRPMQKEILLTCSGHWQNVSLPVD-----CGV-----PDP 1469  
 D 717 ICEENFTMIGHG-----VSCISGKMTQLPVCVATDQLEKCRVLKSTGEAIKP 765  
 QY 1470 SLVNYA-----NFGSCGTYFLKRCST-----SCVPPAKLOGL----- 1502  
 D 766 KLTFTFHTSTMDYKCRDKOYERSICINGKMDPEPNTSKTSCPPQIINTQVIEFTVK 825  
 QY 1503 -----SPWLTCLEDEGLW-SLPVEYCKLECDAPITIANALLPHCL 1542  
 D 826 YLDGEKLSVLCQDNYLTQDSEEMVC-KDGRWQSLPRCTEKIPCSQPTLTHGSLNLRSS 884  
 QY 1543 QD-----NHDVGTICKYECKPGYVAESAGKVANKLKLTQCLEGGIWEQSCIPV 1594  
 D 885 EBRDSTLESSHHGTFPSVVCDDGFRIPREN-----RITCWGKMTSPRCGLP 935  
 QY 1595 CEPPPPV-----FEGMECTNGPSL----- 1614  
 D 936 CGPPPSIPLGTVSLBESYQHGEEVYHCTGFGIDGPAFICGSGKMDPPKICTDCD 995  
 QY 1615 -----DSCGV- 1619  
 D 996 VLPVNAKIIIRGSKSKSYRTGBOVTRCQSPYQMGNSDYTCVNSRMIGQPVCKDNSCVD 1055  
 QY 1620 -----LNCQDERE---KLPILTGKGLTQEFKLCENLOGE 1652

Db 1056 PPHVFNATVTRTKNKLHGDVRVRYECNKPPLFQGVFVWC-ENGWTEKPK-CRDSYTK 1113

Qy 1653 CPPP-----SELNSVEKCEGOGYIG-----VCSPLCV 1682

Db 1114 CGPPPDNGDITSLSPVYPLSSVEYQCKRYLLKSKTITCTNGKMSPPCLHACV 1173

Qy 1683 IPPSDPVALPENITADTLEHMMPEPKVOS 1711

Db 1174 IP--ENIMESHNT-----LKWRTHEKITS 1196

RESULT 7

LEM2 HUMAN STANDARD; PRT; 610 AA.

AC P16581: P16111;

DT 01-APR-1990 (Rel. 14, Created)

DT 01-AUG-1990 (Rel. 15, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE E-selectin precursor (Endothelial leukocyte adhesion molecule 1) (ELAM-1) (Leukocyte-endothelial cell adhesion molecule 2) (LECAM2) (CD62E).

GN SELE OR ELAM1.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

OX NCBI\_TaxID=9606;

[1]

RP SEQUENCE FROM N.A.

RK MEDLINE=90175359; PubMed=1689848;

RA Hession C., Osborn L., Goff D., Chi-Rosso G., Vassallo C., Paasek M., Plickack C., Tizard R., Goelz S., McCarthy K., Hopple S., Lobb R.;

RT "Endothelial leukocyte adhesion molecule 1: direct expression cloning and functional interactions."

RL Proc. Natl. Acad. Sci. U.S.A. 87:1673-1677(1990).

[2]

RP SEQUENCE FROM N.A.

RK MEDLINE=89162047; PubMed=2466335;

RA Bevilacqua M.P., Stengelin S., Gimbrone M.A. Jr., Seed B.;

RT "Endothelial leukocyte adhesion molecule 1: an inducible receptor for neutrophils related to complement regulatory proteins and lectins."

RL Science 243:1160-1165(1989).

[3]

RP SEQUENCE FROM N.A.

RK MEDLINE=91115870; PubMed=1703529;

RA Collins T., Williams A., Johnston G.I., Kim J., Eddy R., Shows T., Gimbrone M.A. Jr., Bevilacqua M.P.;

RT "Structure and chromosomal location of the gene for endothelial-leukocyte adhesion molecule 1."

RL J. Biol. Chem. 266:2466-2473(1991).

[4]

RP SEQUENCE FROM N.A., AND VARIANTS SER-21; ILE-31; ARG-149; PRO-257; LYS-295; GLN-421; TYR-468; SER-550 AND PHE-575.

RK Rider M.J., Armet T.Z., Carrington D.P., Ozuna M., Kuldanek S.A., Rajkumar N., Toth E.J., Yi Q., Mickerson D.A.;

RT Submitted (AUG-2002) to the EMBL/Genbank/DBJ databases.

RL J. Biol. Chem. 266:2466-2473(1991).

[5]

RP SEQUENCE FROM N.A.

RA Pearce A.;

RL Submitted (APR-1998) to the EMBL/Genbank/DBJ databases.

[6]

RP LIGAND.

RK MEDLINE=91068005; PubMed=1701274;

RA Phillips M.L., Nudelmann E., Gaeta F.C., Perez M., Singhal A.K., Hakomori S., Paulson J.C.;

RT "ELAM-1 mediates cell adhesion by recognition of a carbohydrate ligand, sialyl-Lex."

RL Science 250:1130-1132(1990).

[7]

RP 3D-STRUCTURE MODELING OF LECTIN DOMAIN.

RK MEDLINE=93202275; PubMed=7681016;

RA Mills A.;

RT "Modelling the carbohydrate recognition domain of human E-selectin."

RL FEBS Lett. 319:5-11(1993).

[8]

RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 22-178.

RK MEDLINE=94150646; PubMed=7509040;

RA Graves B.J., Crowther R.L., Chandran C., Rumberger J.M., Li S., Huang K.-S., Presky D.H., Familletti P.C., Wollitzky B.A., Burns D.K.;

RT "Insight into E-selectin/ligand interaction from the crystal structure and mutagenesis of the lec/EGF domains."

RL Nature 367:532-538(1994).

[9]

RP VARIANT ARG-149.

RK MEDLINE=95179107; PubMed=7533025;

RA Wenzel K., Felix S., Kleber F.X., Brachold R., Menke T., Schattke S., Schulte K.L., Glaser C., Rohde K., Baumann G., Speer A.;

RT "E-selectin polymorphism and atherosclerosis: an association study."

RL Hum. Mol. Genet. 3:1935-1937(1994).

[10]

RP VARIANTS ARG-149 AND PHE-575.

RK MEDLINE=96140743; PubMed=8557254;

RA Wenzel K., Ernst M., Rohde K., Baumann G., Speer A.;

RT "DNA polymorphisms in adhesion molecule genes -- a new risk factor for early atherosclerosis."

RL Hum. Genet. 97:15-20(1996).

[11]

RP VARIANT ARG-149.

RK MEDLINE=99134508; PubMed=9933738;

RA Ye S.O., Usher D., Virgil D., Zhang L.O., Yochim S.E., Gupta R.;

RT "A beta polymorphism detects the mutation of serine-128 to arginine in CD 62E gene - a risk factor for coronary artery disease."

RL J. Biomed. Sci. 6:18-21(1999).

[12]

RP VARIANTS ARG-149; TYR-468 AND PHE-575.

RK MEDLINE=99318094; PubMed=10391210;

RA Haluska M.K., Fan J.-B., Bentley K., Heie L., Shen N., Weder A., Cooper R., Lipschutz R., Chakravarti A.;

RT "Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis."

RL Nat. Genet. 22:239-247(1999).

[13]

RP VARIANT PHE-575.

RK MEDLINE=20434915; PubMed=10962036;

RA Saes C., Pallaud C., Zannad F., Vissikis S.;

RT "Relationship between E-selectin L/F554 polymorphism and blood pressure in the Stanislas cohort."

RL Hum. Genet. 107:58-61(2000).

- FUNCTION: EXPRESSED ON CYTOKINE INDUCED ENDOTHELIAL CELLS AND MEDIATES THEIR BINDING TO LEUKOCYTES. THE LIGAND RECOGNIZED BY ELAM-1 IS SIALYL-LEWIS X (ALPHA(1-3) FUCCOSYLATED DERIVATIVES OF POLYGLYCOSAMINE THAT ARE FOUND AT THE NONREDUCING TERMINI OF GLYCOLIPIDS).

- SUBCELLULAR LOCATION: Type I membrane protein.

- POLYMORPHISM: A POLYMORPHISM IN POSITION 149 IS ASSOCIATED WITH A HIGHER RISK OF CORONARY ARTERY DISEASE (CAD). A SIGNIFICANTLY HIGHER MUTATION FREQUENCY (ARG-149) IS OBSERVED IN PATIENTS WITH ANGIOGRAPHICALLY PROVEN SEVERE ATHEROSCLEROSIS COMPARED WITH AN UNSELECTED POPULATION (SER-149).

- SIMILARITY: BELONGS TO THE SELECTIN/LECAM FAMILY.

- SIMILARITY: Contains 1 C-type lectin family domain.

- SIMILARITY: Contains 1 EGF-like domain.

- SIMILARITY: Contains 6 Sushi (SCR) domains.

- DATABASE: NAME=PROV; NOTE=CD guide CD62E entry; WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd62e.htm".

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - CC the European Bioinformatics Institute. There are no restrictions on its CC use by non-profit institutions as long as its content is in no way CC modified and this statement is not removed. Usage by and for commercial CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; M30640; AAA52377.1; -

CC EMBL; M61893; AAA52375.1; -

DR	EMBL; M61895; AAA52375.1;	JOINED.
DR	EMBL; M61887; AAA52375.1;	JOINED.
DR	EMBL; M61888; AAA52375.1;	JOINED.
DR	EMBL; M61890; AAA52375.1;	JOINED.
DR	EMBL; M61892; AAA52375.1;	JOINED.
DR	EMBL; M24736; AAA52376.1;	-.
DR	EMBL; AF540378; AA01237.1;	-.
DR	EMBL; AL021940; CAAL7434.1;	-.
DR	PIR; A38615; A35046.	
DR	PDB; 1ESL; 31-AUG-94.	
DR	PDB; 1G1T; 13-OCT-01.	
DR	PDB; 1KJA; 03-APR-96.	
DR	Genew; HGNC:10718; SELE.	
DR	MIM; 131210; -.	
DR	GO; GO:0005886; C:plasma membrane; TNS.	
DR	GO; GO:0006954; P:inflammatory response; TNS.	
DR	Interpro; IPR006209; EGF like.	
DR	Interpro; IPR006210; IEgf.	
DR	Interpro; IPR001304; Lectin C.	
DR	Interpro; IPR002386; Selectin.	
DR	Interpro; IPR000436; Sushi_SCR_CCP.	
DR	Pfam; PF00008; EGF_1.	
DR	Pfam; PF00059; lectin_c; 1.	
DR	Pfam; PF00084; sushi; 6.	
DR	PRINTS; PR00343; SELECTIN.	
DR	SMART; SM00032; CCP; 6.	
DR	SMART; SM00034; CLECT; 1.	
DR	SMART; SM00181; EGF; 1.	
DR	PROSITE; PS00022; EGF_1; 1.	
DR	PROSITE; PS01186; EGF_2; 1.	
DR	PROSITE; PS00615; C TYPE LECTIN_1; 1.	
DR	PROSITE; PSS0041; C TYPE LECTIN_2; 1.	
KM	Cell adhesion; Transmembrane; Glycoprotein; EGF-like domain; Lectin selectin; Signal; Sushi; Repeat; Polymorphism; 3d-structure.	
FT	SIGNAL	1 21
FT	CHAIN	22 610
FT	DOMAIN	22 556
FT	TRANSMEM	557 578
FT	DOMAIN	579 610
FT	DOMAIN	38 138
FT	DOMAIN	139 175
FT	DOMAIN	179 238
FT	DOMAIN	241 300
FT	DOMAIN	303 363
FT	DOMAIN	366 426
FT	DOMAIN	429 489
FT	DOMAIN	492 548
FT	DISULFID	40 138
FT	DISULFID	111 130
FT	DISULFID	143 154
FT	DISULFID	148 163
FT	DISULFID	165 174
FT	DISULFID	180 224
FT	DISULFID	210 237
FT	DISULFID	242 286
FT	DISULFID	272 299
FT	DISULFID	304 349
FT	DISULFID	335 362
FT	DISULFID	367 412
FT	DISULFID	398 425
FT	DISULFID	430 475
FT	DISULFID	461 488
FT	DISULFID	493 534
FT	DISULFID	520 547
FT	CARBOHYD	25 25
FT	CARBOHYD	145 145
		E-SELECTIN.
		EXTRACELLULAR (POTENTIAL).
		POTENTIAL.
		CYTOSOLIC (POTENTIAL).
		C-TYPE LECTIN (SHORT FORM).
		EGF-LIKE.
		SUSHI 1.
		SUSHI 2.
		SUSHI 3.
		SUSHI 4.
		SUSHI 5.
		SUSHI 6.
		BY SIMILARITY.
		BY SIMILARITY.
		BY SIMILARITY.
		BY SIMILARITY.
		BY SIMILARITY.
		BY SIMILARITY.
		BY SIMILARITY.
		BY SIMILARITY.
		BY SIMILARITY.
		N-LINKED (GLCNAC. . .) (POTENTIAL).
		N-LINKED (GLCNAC. . .) (POTENTIAL).

```

Query Match      2.3%, Score 230.5, DB 1, Length 610;
Best Local Similarity 21.1%; Pred. No. 7.3e-08;
Matches 127; Conservative 80; Mismatches 215; Indels 181; Gaps 36;
QY 1230 RPLT-----GMPFCVASENETQDNRSEQEGSLAKKEDV--W-----LKVCNRRPGE 1274

```

Db	88	KPYLEAKNMAP--GEENNRKOD--EDCEVETIYKEKQVGMNDRCSKKLALCY-----	139
Qy	1275	ARAIFILFTDGLVGEHQPFVTLYLTLDVGRSNH-----SLAGTY-----GLSCQH	1320
Db	140	-----TAACNTNNSCSGHGSCVEETINNTYTCKDPDGFGLKCEQ	176
Qy	1321	NPLIINTYTHONVLFHHTTSVLNLFSSPRFGISAVLARTSRIG--LSAPNSC--ISEDEQG	1378
Db	177	---IVNCT-----ALBEPHQ-----SLVCSHPLDNFMYNSSCSISCRGY	214
Qy	1379	NHQQSGCIHRPCGKODSCPSLLDLHADVNNCTSI-----GPGLM-----KCAIT	1422
Db	215	LPSSMETWQCKSSGEMSAF---IPACNVBCDAVTNPANGCFEFGQNPFGSEFPNATTTCTFD	271
Qy	1423	CORGFPALQASGGYIRPMQKEIILITCSSGHD--QNVSCLPYDCGV---PPSLVNYANFS	1478
Db	272	CEEFEL-----MGAGSLQCTSGMMDNKPKPCKAVTCGAVNQPFQNGSVRCHSP	321
Qy	1479	CSBGITKLKRCISICVPAPKLGISPMULTLEDGLSLPEYCK--LECDAPITLNLNL	1537
Db	322	AGEFT--FKSSCNFPCEBGFMLQGPQAQ--VECTQOQMTOQIIPVCEAFOCTA---LSNBERG	376
Qy	1538	LPHL-----QDNHDVGTICKYECKRGYVABABEKVANKLKIKOCLEGGIT--RQSGCIP	1592
Db	377	YMNCLPSASGSGFRYSSCFECSCEGFPVLKSK-----RLQCGPTEGDNKPTCEA	427
Qy	1593	VCPEP--PPVFEBSMYECTNG-----FSLDSQCVLNCQERE---KLPLCTKEGLMTQEF	1643
Db	428	VRCAVAVQPP--KLIVCAHSPIGEFETFKSCASCSEGFELHSTOLECTISQOQWTEV	485
Qy	1644	KLCENLQGECPPEPSBLNS-----VEYKCEQGYIGA---VCS-----P	1679
Db	486	PSQVNVKCSLAVPGKINMSSCGEPVEGTVCKFCAPREGWTLNGSAARTCGATGMSGLLP	545
Qy	1680	LCVTPSPDPVLPENITADTLE--HMMPVKYQGISVCTGR--RQMPDPVLYHCIOSCPEFQ	1737
Db	546	TCEAPTESNIPVLVGLSAGLSLITLTLAPFLMLMKCKIRAKKFPV-----ASSCQSL	598
Qy	1738	ADG 1740	
Db	599	SDG 601	

RESULT 8

DAF2 MOUSE

ID	DAF2 MOUSE	STANDARD;	PRT;	407 AA.
AC	Q61476;			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	01-NOV-1997 (Rel. 35, Last annotation update)			
DE	Complement decay-accelerating factor, transmembrane precursor (DAF-1M).			
DE	DAF2.			
GN	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=Testis;			
RX	MEDLINE=95403982; PubMed=7545711;			
RA	Spicer A.P., Seidin M.F., Gendler S.J.;			
RT	"Molecular cloning and chromosomal localization of the mouse decay-accelerating factor genes. Duplicated genes encode glyco1phosphatidylinositol-anchored and transmembrane forms.";			
RL	J. Immunol. 155:3079-3091(1995).			
CC	-1 FUNCTION: PROTECTION OF CELLS FROM COMPLEMENT-MEDIATED DAMAGE (BY SIMILARITY).			
CC	-1 SUBCELLULAR LOCATION: Type I membrane protein (Potential).			
CC	-1 TISSUE SPECIFICITY: TESTIS, SPLEEN AND LYMPH NODE.			
CC	-1 DOMAIN: THE FIRST SUSHI DOMAIN (SCR1) IS NOT NECESSARY FOR FUNCTION.SCR2 AND SCR4 PROVIDE THE PROPER CONFORMATION FOR THE			

CC ACTIVE SITE ON SCRs (BY SIMILARITY).

CC -1- SIMILARITY: Contains 4 Sushi (SCR) domains.

CC -1- SIMILARITY: BELONGS TO THE RECEPTORS OF COMPLEMENT ACTIVATION (RCA) FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL: L41365; AAB00092.1; -.

CC HSBP; P08603; IHCC.

CC MED; MGI:104849; Daf2.

CC InterPro; IPR000436; Sushi\_SCR\_CCP.

CC Pfam; PF00084; sushi; 4.

CC SMART; SM00032; CCP; 4.

CC Complement pathway; Glycoprotein; Repeat; signal; Sushi; Transmembrane.

CC SIGNAL 1 39 POTENTIAL.

CC CHAIN 40 407 COMPLEMENT DEGRADATION-ACCELERATING FACTOR, TRANSMEMBRANE.

CC DOMAIN 40 368 EXTRACELLULAR (POTENTIAL).

CC TRANSMEM 369 389 POTENTIAL.

CC DOMAIN 390 407 CYTOPLASMIC (POTENTIAL).

CC DOMAIN 40 100 SUSHI 1.

CC DOMAIN 102 164 SUSHI 2.

CC DOMAIN 167 226 SUSHI 3.

CC DOMAIN 229 290 SUSHI 4.

CC DOMAIN 291 363 SER/THR-RICH (BY SIMILARITY).

CC DISULFID 70 99 BY SIMILARITY.

CC DISULFID 103 150 BY SIMILARITY.

CC DISULFID 134 163 BY SIMILARITY.

CC DISULFID 168 209 BY SIMILARITY.

CC DISULFID 195 225 BY SIMILARITY.

CC DISULFID 230 272 BY SIMILARITY.

CC DISULFID 258 289 BY SIMILARITY.

CC CARBOHYD 192 192 N-LINKED (GLCNAC. . .) (POTENTIAL).

CC CARBOHYD 267 267 N-LINKED (GLCNAC. . .) (POTENTIAL).

CC SEQUENCE 407 AA; 44469 MW; DDD82P2D72CEAD40 CRC64;

SO

Query Match 2.3%; Score 229; DB 1; Length 407;

Best Local Similarity 27.2%; Pred. No. 5.1e-08;

Matches 71; Conservative 31; Mismatches 101; Indels 58; Gaps 10;

Qy 1463 DCGVDP-----PSLVNYANF-----SCSEGTGFLKRCISICVPPAKUQGLSPMLTJC 1508

Db 40 DCGPPDIPNARPIGRHSKFAEGSKVAVSCNNGPK-----GVPMKSNIVVC 86

Qy 1509 LEDGMSLPYVYCKLECAPPIILNANILLPHCLDNHVDGTCYCKEKKPGYVAESAG 1568

Db 87 LENGQMSHEHTEFCESKCTPERLSFKKEVFNNAPFVGVYVECRPGRKOPSLSG 146

Qy 1569 KVRNKLTLIOCLLEGSIWEGGS--CIPIVCEPPPEVFEAGYEECTNGFSLDSOCVLANCNER 1626

Db 147 -----KSTCLIEDLVMSPVAFCKKSCPNPKDLDNHINIPITILGSEINNSCANGY 199

Qy 1627 EKL-----PICTKEGL--WTOEFKLCENIQCCEPPP-----SEINL-----VEYK 1666

Db 200 RLVGITSLITCTITGNAVMDDEFVCTEI--FCPPPKIKDGIKMGESDSYKVSQVVIYS 257

Qy 1667 CEQGIGIAGVCSPLCVIPPSD 1687

Db 258 CDKGIFLFGNSTIYCTVSKSD 278

RESULT 9

LEM3\_RAT

ID LEM3\_RAT STANDARD; PRT; 768 AA.

AC P98106;

DT 01-FEB-1996 (Rel. 33, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE P-selectin precursor (Granule membrane protein 140) (GMP-140) (PADGEM) (CD62P) (leukocyte-endothelial cell adhesion molecule 3) (LECAM3).

GN SELP.

OS Rattus norvegicus (Rat).

OC Eukaryote; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Euteheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI\_Taxid=10116;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Lung;

RX MEDLINE=94333817; PubMed=7520013;

RA Auchampach J.A., Oliver M.G., Anderson D.C., Manning A.M.;

RT "Cloning, sequence comparison and in vivo expression of the gene encoding rat P-selectin."

RL Gene 145:251-255(1994).

CC -1- FUNCTION: CA(2+)-DEPENDENT RECEPTOR FOR MYELOID CELLS THAT BINDS TO CARBOHYDRATES ON NEUTROPHILS AND MONOCYTES. MEDIATES THE INTERACTION OF ACTIVATED ENDOTHELIAL CELLS OR PLATELETS WITH LEUKOCYTES. THE LIGAND RECOGNIZED IS SIATYL-LEWIS X.

CC -1- SUBCELLULAR LOCATION: Type I membrane protein.

CC -1- TISSUE SPECIFICITY: EXPRESSED IN ALL TISSUES EXAMINED: SPLEEN, LUNG, BRAIN, LIVER, HEART, KIDNEY, THYMUS, SMALL INTESTINE.

CC -1- INDUCTION: BY ACUTE INFLAMMATION (PROBABLE).

CC -1- SIMILARITY: BELONGS TO THE SELECTIN/LECAM FAMILY.

CC -1- SIMILARITY: Contains 1 C-type lectin family domain.

CC -1- SIMILARITY: Contains 1 EGF-like domain.

CC -1- SIMILARITY: Contains 8 Sushi (SCR) domains.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL: L23088; AAA60325.1; -.

CC PIR; I53821; I53821.

CC HSBP; P16109; IFSB.

CC InterPro; IPR006209; EGF\_1like.

CC InterPro; IPR001304; Lectin\_C.

CC InterPro; IPR000436; Sushi\_SCR\_CCP.

CC Pfam; PF00059; lectin\_C; 1.

CC Pfam; PF00084; sushi; 8.

CC SMART; SM00032; CCP; 8.

CC PROSITE; SM00034; CLECT; 1.

CC PROSITE; PS00022; EGF\_1; 1.

CC PROSITE; PS01186; EGF\_2; 1.

CC PROSITE; PS00615; C-TYPE LECTIN\_1; 1.

CC PROSITE; PS50041; C-TYPE LECTIN\_2; 1.

CC Cell adhesion; Transmembrane; Glycoprotein; EGF-like domain; Lectin; Selectin; Signal; Sushi; Repeat; Lipoprotein; Palmitate.

CC SIGNAL 1 41 POTENTIAL.

CC CHAIN 42 768 P-SELECTIN.

CC TRANSMEM 42 709 EXTRACELLULAR (POTENTIAL).

CC DOMAIN 710 733 POTENTIAL.

CC TRANSMEM 734 733 CYTOPLASMIC (POTENTIAL).

CC DOMAIN 58 158 C-TYPE LECTIN (SHORT FORM).

CC DOMAIN 159 195 EGF-LIKE.

CC DOMAIN 199 258 SUSHI 1.

CC DOMAIN 261 320 SUSHI 2.

CC DOMAIN 323 382 SUSHI 3.

CC DOMAIN 385 444 SUSHI 4.

CC DOMAIN 447 506 SUSHI 5.

CC DOMAIN 509 568 SUSHI 6.

CC DOMAIN 579 638 SUSHI 7.

CC DOMAIN 641 700 SUSHI 8.

CC DISULFID 60 158 BY SIMILARITY.

CC DISULFID 131 150 BY SIMILARITY.

CC DISULFID 168 183 BY SIMILARITY.

CC DISULFID 185 194 BY SIMILARITY.

FT DISULFID 200 244 BY SIMILARITY.  
 FT DISULFID 230 257 BY SIMILARITY.  
 FT DISULFID 262 306 BY SIMILARITY.  
 FT DISULFID 292 319 BY SIMILARITY.  
 FT DISULFID 324 368 BY SIMILARITY.  
 FT DISULFID 354 381 BY SIMILARITY.  
 FT DISULFID 386 430 BY SIMILARITY.  
 FT DISULFID 416 443 BY SIMILARITY.  
 FT DISULFID 448 492 BY SIMILARITY.  
 FT DISULFID 478 505 BY SIMILARITY.  
 FT DISULFID 510 554 BY SIMILARITY.  
 FT DISULFID 540 567 BY SIMILARITY.  
 FT DISULFID 580 624 BY SIMILARITY.  
 FT DISULFID 610 637 BY SIMILARITY.  
 FT DISULFID 642 686 BY SIMILARITY.  
 FT DISULFID 672 699 BY SIMILARITY.  
 FT CARBOHYD 45 45 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 54 54 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 107 107 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 212 212 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 347 347 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 456 456 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 603 603 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 654 654 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 661 661 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 679 679 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT LIPID 745 745 PALMITATE (BY SIMILARITY).  
 FT SITE 756 759 ENDOCYTOSIS SIGNAL (PROBABLE).  
 SQ SEQUENCE 768 AA; 83517 MW; 26FD7B8A5F31316 CRC64;

Query Match 2.3%; Score 227.5; DB 1; Length 768;  
 Best Local Similarity 20.4%; Pred. No. 1.6e-07;  
 Matches 105; Conservative 67; Mismatches 176; Indels 167; Gaps 28;

QY 1365 LSAPNSCTSEDEGON-HQGSGCH-----RPGKOD-----SCPSLLLD 1402  
 DB 110 LTAEAEMWADNEPNNKNNQDCEIYIKNSGAPKWNDEPFKRRKALCYTASQDM--- 166  
 QY 1403 HADVNTCTSTGP-----GLMKCAITCGRGFALQSSQYIRPMQKEILLTSSGHWONV 1457  
 DB 167 -----SCNSQGERIETIGSYTC--SCYPGF--YEPCEGYV----- 198  
 QY 1458 SCLPVDGCG---VPPSLVNVYANFSCSE---GTFELKRCISICVPPAKLQGLSPWLTCLD 1511  
 DB 199 -----EGCKPDIPGHVLMN-----CSHPLGSPSSSQTFPCPGYDLNGSE--MQLLAS 247  
 QY 1512 GLWSLPEVYCK-LBCDAPPIILNANLLPHCLQDNHVGITCKYECKRGYVVAESAGKV 1570  
 DB 248 GIWTTNPPQCKAVQCSLEAPLHGTMDCTHPLA-AFAYDSSCKFCQPGY-----RM 298  
 QY 1571 RNKLKIQCLEGGIWEQ--GSCIVVCEPPPPVPEGMEYC---TNGSLDQCYLNCNQE 1625  
 DB 299 RGSDI-LHCTDSQGMSEPLPTCEAIACEPLSPHSGSDCPSPSGAGYNSCTFRCTEG 357  
 QY 1626 REKL---PICTCKGLMTQEFKLCENLOGECPPEPSEL-----NSVEYKCE 1668  
 DB 358 FVLMGNDAIHICADIGQRTAPRVCBALQCCFPPPSKAQVSCSPFGPLKQASHCSFSCD 417  
 QY 1669 QG-----YGIAGVCSPLCVIPSPDPVLPENTITADTLE----- 1701  
 DB 418 EGSILVAGSVIRCLATGHMSBAPEQCAVSCPTLSPENGIMTICLOPLGHSNYKSTCOFM 477  
 QY 1702 -----HMEPEKYVSIVCTGRQWHPDVL---VHCIOSCPFPQADGKCDITNNR-----A 1749  
 DB 478 CDEGFYLSGPERLD---CSPSGHWTGSPMCEAIKCPETFAPEQGLSDCSHVHGEFSVGS 534  
 QY 1750 YCHYDGGDCS---STLSKKYIPFPAADCDDECT 1781  
 DB 535 TCHFS-----CNEFEELGSRNV-----ECT 555

RESULT 10  
 LEM2\_RABIT

ID LEM2\_RABIT STANDARD; PRT; 551 AA.  
 AC P27113;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE E-selectin precursor (Endothelial leukocyte adhesion molecule 1)  
 DE (ELAM-1) (Leukocyte-endothelial cell adhesion molecule 2) (LECAM2)  
 DE (CD62E).  
 GN SELE.  
 OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 OX NCBI\_TaxID=9986;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Heart;  
 RX MEDLINE=92189729; PubMed=1372169;  
 RA Lalligan J.D., Teang T.C., Rumberger J.M., Burns D.K.;  
 RT "Characterization of cDNA and genomic sequences encoding rabbit  
 RT ELAM-1: conservation of structure and functional interactions with  
 RT leukocytes".  
 RL DNA Cell Biol. 11:149-162(1992).  
 CC -1- FUNCTION: EXPRESSED ON CYTOKINE INDUCED ENDOTHELIAL CELLS AND  
 CC MEDIATES THEIR BINDING TO LEUKOCYTES. THE LIGAND RECOGNIZED BY  
 CC ELAM-1 IS STALYL-LEWIS X (ALPHA(1-3)FUCOSYLATED DERIVATIVES OF  
 CC POLYLACTOSAMINE THAT ARE FOUND AT THE NONREDUCING TERMINI OF  
 CC GLYCOLIPIDS).  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- INDUCTION: BY CYTOKINES.  
 CC -1- SIMILARITY: BELONGS TO THE SELECTIN/LECAM FAMILY.  
 CC -1- SIMILARITY: Contains 1 C-type lectin family domain.  
 CC -1- SIMILARITY: Contains 1 EGF-like domain.  
 CC -1- SIMILARITY: Contains 5 Sushi (SCR) domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M91004; AAA31243.1; -  
 DR EMBL; M91005; AAA31244.1; -  
 DR PIR; I46709; I46709.  
 DR HSSP; P16581; IKCA.  
 DR InterPro; IPR006209; EGF-like.  
 DR InterPro; IPR006210; IEGF.  
 DR InterPro; IPR001304; Lectin C.  
 DR InterPro; IPR002396; Selectin.  
 DR InterPro; IPR000436; Sushi\_SCR\_CCP.  
 DR Pfam; PF00008; EGF\_1.  
 DR Pfam; PF00059; lectin\_C\_1.  
 DR Pfam; PF00084; sushi\_5.  
 DR PRINTS; PR00343; SELECTIN.  
 DR SMART; SM00032; CCP\_5.  
 DR SMART; SM00034; CLECT\_1.  
 DR SMART; SM00181; EGF\_1.  
 DR PROSITE; PS00022; EGF\_1; 1.  
 DR PROSITE; PS01186; EGF\_2; 1.  
 DR PROSITE; PS00615; C-TYPE LECTIN\_1; 1.  
 DR PROSITE; PS50041; C-TYPE LECTIN\_2; 1.  
 DR Cell adhesion; Transmembrane; Glycoprotein; EGF-like domain; Lectin;  
 DR Selectin; Signal; Sushi; Repeat.  
 KM SIGNAL 1 23  
 FT CHAIN 24 551  
 FT DOMAIN 24 495  
 FT TRANSMEM 496 517  
 FT DOMAIN 518 551  
 FT DOMAIN 40 140  
 FT DOMAIN 141 177  
 FT DOMAIN 181 240  
 FT DOMAIN 243 302  
 E-SELECTIN.  
 EXTRACELLULAR (POTENTIAL).  
 POTENTIAL.  
 CYTOPLASMIC (POTENTIAL).  
 C-TYPE LECTIN (SHORT FORM).  
 EGF-LIKE.  
 SUSHI 1.  
 SUSHI 2.

```

FT DOMAIN 305 365 SUSHI 3.
FT DOMAIN 368 428 SUSHI 4.
FT DOMAIN 431 487 SUSHI 5.
FT DISULFID 42 140 BY SIMILARITY.
FT DISULFID 113 132 BY SIMILARITY.
FT DISULFID 145 156 BY SIMILARITY.
FT DISULFID 150 165 BY SIMILARITY.
FT DISULFID 167 176 BY SIMILARITY.
FT DISULFID 182 226 BY SIMILARITY.
FT DISULFID 212 239 BY SIMILARITY.
FT DISULFID 244 288 BY SIMILARITY.
FT DISULFID 274 301 BY SIMILARITY.
FT DISULFID 306 351 BY SIMILARITY.
FT DISULFID 337 364 BY SIMILARITY.
FT DISULFID 369 414 BY SIMILARITY.
FT DISULFID 400 427 BY SIMILARITY.
FT DISULFID 432 473 BY SIMILARITY.
FT DISULFID 459 486 BY SIMILARITY.
FT CARBOHYD 32 32 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 45 45 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 162 162 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 194 194 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 201 201 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 205 205 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 267 267 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 314 314 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 321 321 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 334 334 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 442 442 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 466 466 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 308 308 T -> A (IN REF. 1; AAA31244).
FT CONFLICT 328 328 T -> V (IN REF. 1; AAA31244).
FT CONFLICT 491 491 A -> V (IN REF. 1; AAA31244).
SQ SEQUENCE 551 AA; 60346 MW; 23BC8A83B23240E CRC64;

```

```

Query Match 2.3%; Score 225; DB 1; Length 551;
Best Local Similarity 19.8%; Pred. No. 1.5e-07;
Matches 104; Conservative 61; Mismatches 179; Indels 180; Gaps 26;

```

```

QY 1257 LKEDVWLKCYCFNRPRGARAFIFLTLDGL--VRGE-----HQPVVITL----- 1301
DB 76 IRKXNVNVIWGTNHP-----LLEGAKNNAFGSPNNKQNNEDVEIYIKRPKDTG 125
QY 1302 --TVRGSNHSIG--TYGLSCQHNPLIINVTHQNVLFHTTSSVLLNFSPRGISAVALR 1358
DB 126 MANDERCSKKTALCTAAC----- 145
QY 1359 TSSRIGLAPNSCISEDEGQNHQGSCTHRP-----CGKODSCPSLLLDHADVNVCT-S 1411
DB 146 --TEASCSGHGECI-----ETINNSCKCYPGFSGLKCEQVYVTCGAQVQPHGSLNCTHP 198
QY 1412 IG--PGLMKCAITCORGFALQASSQYIRPMQKEILLTCSSGHD--QVNSCLPDC--G 1465
DB 199 LGNSYSSSCSVSCRGYL-----PSTETWTCTTSGSEWAPATCKVVECDTNG 248
QY 1466 VPDESLVNYANFSCSEGTFRFLKRCISCVPRKLOGLSPM-----L 1506
DB 249 KP-----AN-----GVKCSFS--QGSAPMNTTCTPDEBEGFTLLGASL 286
QY 1507 TCLEGLWSLPEVYCK--LECDAPPIILNANILLPHCLDNDHVGITCYECKPGYVAES 1565
DB 287 OCTSGSVDNEKPTCAVSCDPIIHHPQGSVSCNSSGKFTFRSCNFTCEBNPLNGP 346
QY 1566 AEGKRNKLKIQCLEGGIWEQGS--CIPVCEPPPPPEKMEYECTN-----GFSLDSQCV 1619
DB 347 AQ-----VECTAQQQMTQDAVCAVVKDPHPLTLEDGVKCTHHPHTEFFYKSSCT 397
QY 1620 LINCNGERS--KLPICTKREGMTQEPFLCENLOGECPPEPSELNVSAYKCEQGVIGAV 1676
DB 398 FNCREGEFLHSAQULECTISQGMADLPSQVQ--C-PSLAVLAKTNVSCSGEPVFTV 454
QY 1677 CSPLCVIPPSPDVMVLPENITADTLTHMMEPVAVQSIIVCTGRROW 1720

```

```

DB 455 CNFAC-----PEGWTLN-----GSAALMCAEGCOM 479
RESULT 11
ID LTRBS_HUMAN STANDARD; PRT; 1394 AA.
AC P22064; O8TD95;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Latent transforming growth factor beta binding protein, isoform 15
DE precursor (LTBP-1) (Transforming growth factor beta-1 binding protein
DE 1) (TGF-beta1-BP-1).
GN LTBP1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
(1)
SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Fibroblast, and Platelet;
RX MEDLINE=90275601; PubMed=2350783;
RA Kanzaki T., Olofsson A., Moren A., Wernstedt C., Hellman U.,
RA Miyazono K., Claesson-Welsh L., Heldin C.-H.;
RT "TGF-beta 1 binding protein: a component of the large latent complex
RT of TGF-beta 1 with multiple repeat sequences.";
RL Cell 61:1051-1061(1990).
(2)
SEQUENCE FROM N.A.
RA Kwak J.H., Shin K.Y., Kim S.I.;
RT "Major alternative spliced-form of LTBP1 mRNA in human glomerular
RT endothelial cell.";
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RP INTERACTION WITH FIBRILLIN.
RX PubMed=12429738;
RA Isogai Z., Ouo R.N., Uehiro S., Keene D.R., Chen Y., Mazzieri R.,
RA Charbonneau N.L., Reinhardt D.P., Rifkin D.B., Sakai L.Y.;
RT "Latent transforming growth factor beta-binding protein 1 interacts
RT with fibrillin and is a microfibril-associated protein.";
RL J. Biol. Chem. 278:2750-2757(2003).
CC -1 SUBUNIT: The large latent complex of TGF-beta1 from platelets is
CC composed of the TGF-beta1 molecule noncovalently associated with a
CC disulfide-bonded complex of a dimer of the N-terminal propeptide
CC of the TGF-beta1 precursor and a third component denoted TGF-
CC beta1-BP. TGF-beta1-BP does not bind directly to active TGF-BETA1.
CC Binds to fibrillin.
CC -1 SUBCELLULAR LOCATION: Secreted.
CC -1 ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=short;
CC IsoId=P22064-1; Sequence=Displayed;
CC Name=long;
CC IsoId=Q14766-1; Sequence=External;
CC -1 PM: CONTAINS HYDROXYLATED ASPARAGINE RESIDUES.
CC -1 PM: The N-terminus is blocked.
CC -1 SIMILARITY: Contains 16 EGF-like domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M34057; AAA61160.1; -
DR EMBL; AF489528; AAM03124.1; -
DR PIR; A35626; A35626.
DR HSSP; P00750; TTPG.
DR GlycositedB; P22064; -
DR Genew; HGNC:6714; LTBP1.
DR MIM; 150390; -

```



```

QY 1390 -----CGKODSCPSLLDADVYNCISI-----GPIAMKALTCORGF 1427
D 569 YTCICIEYRESEQRKC-----VIDEICTYOHLCOSRCENTBESFLC--ICPAGF 619
QY 1428 ALQASSG-----QYRPMQKEILLTSSGHMDONT-----1457
D 620 -MASEGNCIDVDECLRPD-----VCGEGHCNVTVGAFRCCEYCDSGYRMTGRCEDI 672
QY 1458 -SCU-PVDCGVPDSLVVYVYANFSCSEGTKFLKRSISCVPAKIQGLSPML--TCLBGLM 1514
D 673 DECNLPSTC--PDQCVN-----SPG-----SYQCVPT--EGFRGMGQCLDNDV-- 713
QY 1515 SLPEVYCTGECDAPIILN--ANLL-----LPHCLDNHNVG 1549
D 714 -----ECLERPVCANGDCSNLESGYMSCHKGYTRTDHAKHCRDIDCQGNLCVN 764
QY 1550 TICK-----YECKPGYVAES-----AEKVRNKLKIQCL-----1580
D 765 GQCKNTGSPRCTGQGYQLSAKDQCEIDECQHRHLCAHQCRNTGSPQCVCDQGYR 824
QY 1581 -----EGTIEGSGSCIPVCEPPPVVEGMYECT--NGESLDS--QC 1618
D 825 ASGLGDHCEIDNECLEDKSVQCRGDCINTA-----GSYDCTCPDGFQIDNKTQD 875
QY 1619 VLNCGQERKLPILCTKEGLWTQEPKLCENTQCECPPELSNVEYKCEQGYGI---GA 1675
D 876 INECEH-----PGLCGPQG-----ECLNTEG-----SFHCVCQGSFISADGR 913
QY 1676 VCSPL--CVIPSPSPVMLPENITADTLEHMEPVKVOSIVCTGRORHPDVLVHCTQSC 1733
D 914 TCEDIDECV--NNTVCDSS--HGFCMDTAGSFRL-----CYGGF 948
QY 1734 EPPQADGKCDITN-----NRAYC-HYDGGCCSSITLSKVIYIPPAAC-----DL 1777
D 949 QAPDGGQCVVNVCECELLSVCGEAFCEVNEGFLVCACADENQESYPMTGQCRSTSTDL 1008
QY 1778 DECTCRDPKAEENQ 1791
D 1009 D-VVDQPKERKE 1021

RESULT 12
DAFL MOUSE STANDARD. PRT: 390 AA.
AC Q61475, P87732, Q61397,
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Complement decay-accelerating factor, GPI-anchored precursor
DE (DAF-GPI).
GN DAF1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=95403982; PubMed=745711;
RA Spicer A.P., Seldin M.F., Gendler S.J.;
RT "Molecular cloning and chromosomal localization of the mouse decay-
RT accelerating factor genes. Duplicated genes encode
RT glycosylphosphatidylinositol-anchored and transmembrane forms.";
RL J. Immunol. 155:3079-3091 (1995).
RN [2]
RP SEQUENCE OF 7-390 FROM N.A.
RC STRAIN=BALB/c; TISSUE=Spleen;
RX MEDLINE=96362213; PubMed=8671624;
RA Fukunaka Y., Yasui A., Okada N., Okada H.;
RT "Molecular cloning of murine decay accelerating factor by
RT immunoscreening.";
RL Int. Immunol. 8:379-385 (1996).

```

```

CC -1- FUNCTION: PROTECTION OF CELLS FROM COMPLEMENT-MEDIATED DAMAGE (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor (By
CC similarity).
CC -1- TISSUE SPECIFICITY: BRAIN, SECRETORY EPITHELIA, SKELETAL MUSCLE,
CC LIVER, TESTES, THYMUS, SPLEEN AND LYMPH NODE.
CC -1- DOMAIN: THE FIRST SUSHI DOMAIN (SCR1) IS NOT NECESSARY FOR THE
CC FUNCTION. SCR2 AND SCR4 PROVIDE THE PROPER CONFORMATION FOR THE
CC ACTIVE SITE ON SCR3 (BY SIMILARITY).
CC -1- SIMILARITY: Contains 4 Sush1 (SCR) domains.
CC -1- SIMILARITY: BELONGS TO THE RECEPTORS OF COMPLEMENT ACTIVATION
CC (RCA) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L41366; BAB00091.1; -.
DR EMBL; D63679; BAA09830.1; -.
DR HSSP; P08603; IHCC.
DR MCD; MGI:104850; Dafi.
DR InterPro; IPR000436; Sush1_SCR_CCP.
DR Pfam; PF00094; sush1; 4.
DR SMART; SM00032; CCP; 4.
DR KMW Complement pathway; Glycopoltein; Membrane; Repeat; GPI-anchor;
DR Signal; Sush1.
FT FT CHAIN 1 34
FT FT SIGNAL 35 362
FT FT PROPEP 363 390
FT FT DOMAIN 35 95
FT FT DOMAIN 97 159
FT FT DOMAIN 162 221
FT FT DOMAIN 224 285
FT FT DOMAIN 288 362
FT FT DISULFID 65 94
FT FT DISULFID 98 145
FT FT DISULFID 129 158
FT FT DISULFID 163 204
FT FT DISULFID 190 220
FT FT DISULFID 225 267
FT FT DISULFID 253 284
FT FT CARBOHYD 187 187
FT FT CARBOHYD 262 262
FT FT LIPID 362 362
FT FT CONFLICT 7 7
FT FT CONFLICT 9 9
FT FT CONFLICT 83 83
FT FT CONFLICT 91 91
FT FT CONFLICT 135 135
FT FT CONFLICT 173 173
FT FT CONFLICT 180 180
SQ SEQUENCE 390 AA; 42618 MW; 4418721DFF47F8E7 CRC64;

Query Match 2.38; Score 223; DB 1; Length 390;
Best local similarity 25.78; Pred. No. 1.2e-07;
Matches 67; Conservative 31; Mismatches 105; Indels 58; Gaps 9;

QY 1463 DCGPPD-----PELVVYANP-----SCSEGTKFLKRSISCVPAKIQGLSPMLTC 1508
D 35 DCGPPDIPNARPLTGHRSKFAEQSKVAYSCNNGFK-----QVPDKSNIVVC 81
QY 1509 LEDLMSLPEVYCTGECDAPIILNANLLPHCLDNHNVGTICKYCKEYVAESAEG 1568
D 82 LENQMSHHEFFCKSCVAPERLSFASLKEEYLMNFFPGTIVYECRCGRFQRPPLPG 141
QY 1569 KVRNKLKIQCLGEGWEGS--CIPVCEPPPVVEGMYECTNGSLDSQCVLNCQER 1626
D 142 KA-----TCLBGLVMSVPAQFCCKKSCPNKOLDNGHINIPGLIFGSEINFSNPGY 194

```



QY 1627 EKLPI---LCTKGL---WTQEFKLCENLQCECPPSELSN-----VEYK 1666  
 DB 195 RLTVGASTFCSVTGNVTWMDPEFVCTEI--HCEPEPKINNGIMRGSDSYTSQVYTS 252  
 QY 1667 CEGCYGAGNCSPICVPIPPSD 1687  
 DB 253 CDKGFILVGNASIVCTVSKSD 273

RESULT 13  
 DAF\_HUMAN STANDARD; PRT; 381 AA.  
 ID DAF\_HUMAN P08174; P08679; P78361;  
 AC P08174; P08679; P78361;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Complement decay-accelerating factor precursor (CD55 antigen).  
 GN DAF OR CR CD55.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxId=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
 RX MEDLINE=87115845; PubMed=2433596;  
 RA Carae I.W., Davitz M.A., Rhee L., Weddell G., Martin D.W. Jr.,  
 RA Nussenzweig V.;  
 RT "Cloning of decay-accelerating factor suggests novel use of splicing  
 RT to generate two proteins.";  
 RL Nature 325:545-549(1987).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX TISSUE=Cervix;  
 MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.H., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Samperton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loughran N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Huljk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Heaton E., Kettlemann M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Matra M.A.,  
 RA "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [3]  
 RP SEQUENCE OF 6-381 FROM N.A. (ISOFORM 2).  
 RX MEDLINE=87175602; PubMed=2436222;  
 RA Medof M.E., Lublin D.M., Holers V.M., Ayers D.J., Getty R.R.,  
 RA Leykam U.F., Atkinson J.P., Tykocinski M.L.;  
 RT "Cloning and characterization of cDNAs encoding the complete sequence  
 RT of decay-accelerating factor of human complement.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 84:2007-2011(1987).  
 RN [4]  
 RP SEQUENCE OF 35-381 FROM N.A. (ISOFORM 2).  
 RX TISSUE=Hippocampus;  
 RA Kumar V.B., Hyung C., Nakra R., Walters M., Sasser T., Bernardo A.;  
 RT "Decay-accelerating factor (DAF; CD 55) in the brain of Alzheimer's  
 RT disease patients.";  
 RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE OF 1-100 FROM N.A.  
 RX MEDLINE=91271256; PubMed=1711208;

RA Ewulonu U.K., Ravi L., Medof M.E.;  
 RT "Characterization of the decay-accelerating factor gene promoter  
 RT region.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 88:4675-4679(1991).  
 RN [6]  
 RP SEQUENCE OF 35-46.  
 RX TISSUE=Urine;  
 MEDLINE=91291869; PubMed=1712233;  
 RA Nakano Y., Sugita Y., Ishikawa Y., Choi N.-H., Tobe T., Tomita M.;  
 RT "Isolation of two forms of decay-accelerating factor (DAF) from human  
 RT urine.";  
 RL Biochim. Biophys. Acta 1074:326-330(1991).  
 RN [7]  
 RP GPI-ANCHOR.  
 RX MEDLINE=91093238; PubMed=1824699;  
 RA Moran P., Raab H., Kohr W.J., Carae I.W.;  
 RT "Glycophospholipid membrane anchor attachment. Molecular analysis of  
 RT the cleavage/attachment site.";  
 RL J. Biol. Chem. 266:1250-1257(1991).  
 RN [8]  
 RP DISULFIDE BONDS IN SUSHI DOMAINS.  
 RX MEDLINE=92305034; PubMed=1377029;  
 RA Nakano Y., Sumida K., Kikuta N., Mura N.-H., Tobe T., Tomita M.;  
 RT "Complete determination of disulfide bonds localized within the short  
 RT consensus repeat units of decay accelerating factor (CD55 antigen).";  
 RL Biochim. Biophys. Acta 1116:235-240(1992).  
 RN [9]  
 RP FUNCTION AS A ECHOVIRUS RECEPTOR.  
 RX MEDLINE=95045399; PubMed=7525274;  
 RA Ward T., Pipkin P.A., Clarkson N.A., Stone D.M., Minor P.D.,  
 RA Almond J.W.;  
 RT "Decay-accelerating factor CD55 is identified as the receptor for  
 RT echovirus 7 using CELFCS, a rapid immuno-focal cloning method.";  
 RL EMBO J. 13:5070-5074(1994).  
 RN [10]  
 RP VARIANT BLOOD GROUP DR(A-).  
 RX MEDLINE=94325573; PubMed=7519480;  
 RA Lublin D.M., Mallinson G., Poole J., Reid M.E., Thompson E.S.,  
 RA Fardman B.R., Telen M.J., Anstee D.J., Tanner M.J.A.;  
 RT "Molecular basis of reduced or absent expression of  
 RT decay-accelerating factor in Cromer blood group phenotypes.";  
 RL Blood 84:1276-1282(1994).  
 RN [11]  
 RP FUNCTION: THIS PROTEIN RECOGNIZES C4B AND C3B FRAGMENTS THAT  
 CC CONDENSE WITH CELL-SURFACE HYDROXYL OR AMINO GROUPS WHEN NASCENT  
 CC C4B AND C3B ARE LOCALLY GENERATED DURING C4 AND C3 ACTIVATION  
 CC INTERFERES WITH THEIR ABILITY TO CATALYZE THE CONVERSION OF C2 AND  
 CC FACTOR B TO ENZYMATICALLY ACTIVE C2A AND BB AND THEREBY PREVENTS  
 CC THE FORMATION OF C4B2A AND C3BBB, THE AMPLIFICATION CONVERTASES OF  
 CC THE COMPLEMENT CASCADE.  
 CC -1- FUNCTION: ALSO ACTS AS THE RECEPTOR FOR ECHOVIRUS 7 AND RELATED  
 CC VIRUSES (ECHOVIRUSES 13, 21, 29 AND 33).  
 CC -1- SUBUNIT: MONOMER (MAJOR FORM) AND NON-DISULFIDE-LINKED, COVALENT  
 CC HOMODIMER (MINOR FORM).  
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=2; Synonyms=DAF-2;  
 CC IsoId=P08174-1; Sequence=Displayed;  
 CC Name=1; Synonyms=DAF-1;  
 CC IsoId=P08174-2; Sequence=VSP 001200;  
 CC TISSUE SPECIFICITY: EXPRESSED ON THE PLASMA MEMBRANES OF ALL CELL  
 CC TYPES THAT ARE IN INTIMATE CONTACT WITH PLASMA MEMBRANE  
 CC PROTEINS. IT IS ALSO FOUND ON THE SURFACES OF EPITHELIAL CELLS  
 CC LINING EXTRACELLULAR COMPARTMENTS, AND VARIANTS OF THE MOLECULE  
 CC ARE PRESENT IN BODY FLUIDS AND IN EXTRACELLULAR MATRIX.  
 CC -1- DOMAIN: THE FIRST SUSHI DOMAIN (SCR1) IS NOT NECESSARY FOR  
 CC FUNCTION. SCR2 AND SCR4 PROVIDE THE PROPER CONFORMATION FOR THE  
 CC ACTIVE SITE ON SCR3 (BY SIMILARITY).  
 CC -1- PTM: THE SER/THR-RICH DOMAIN IS HEAVILY O-GLYCOSYLATED.  
 CC -1- POLYMORPHISM: DAF IS RESPONSIBLE FOR THE CROMER BLOOD GROUP  
 CC SYSTEM. IT CONSISTS OF AT LEAST SEVEN HIGH-INCIDENCE (CR(A),  
 CC TC(A), DR(A), ES(A), WES(B), UMC, AND IFC) AND LOW-INCIDENCE



DR HSSP; P16109; 1FSB.  
 DR MGD; MGI:98280; Selp.  
 DR GO; GO:0006954; P:inflammatory response; IMP.  
 DR Interpro; IPR000742; EGF\_2.  
 DR Interpro; IPR006209; EGF\_1like.  
 DR Interpro; IPR006210; IEGF.  
 DR Interpro; IPR001304; Lectin C.  
 DR Interpro; IPR002366; Selectin.  
 DR Interpro; IPR000436; Sushi\_SCR\_CCP.  
 DR Pfam; PF00008; EGF\_1.  
 DR Pfam; PF00059; lectin\_c; 1.  
 DR Pfam; PF00084; sushi; 8.  
 DR PRINTS; PR00343; SELECTIN.  
 DR SMART; SM00032; CCP; 8.  
 DR SMART; SM00034; CLECT; 1.  
 DR SMART; SM00181; EGF; 1.  
 DR PROSITE; PS00022; EGF\_1; 1.  
 DR PROSITE; PS01186; EGF\_2; 1.  
 DR PROSITE; PS00615; C-TYPE LECTIN\_1; 1.  
 DR PROSITE; PS50041; C-TYPE LECTIN\_2; 1.  
 DR Cell adhesion; Transmembrane; Glycoprotein; EGF-like domain; Lectin;  
 KW Selectin; Signal; Sushi; Repeat; Lipoprotein; Palmitate.  
 FT SIGNAL 1 41  
 FT CHAIN 42 768  
 FT DOMAIN 42 709  
 FT TRANSSEM 710 733  
 FT DOMAIN 734 768  
 FT DOMAIN 58 158  
 FT DOMAIN 159 195  
 FT DOMAIN 199 258  
 FT DOMAIN 261 320  
 FT DOMAIN 323 382  
 FT DOMAIN 385 444  
 FT DOMAIN 447 506  
 FT DOMAIN 509 568  
 FT DOMAIN 579 638  
 FT DOMAIN 641 700  
 FT DISULFID 60 158  
 FT DISULFID 131 150  
 FT DISULFID 163 174  
 FT DISULFID 168 183  
 FT DISULFID 185 194  
 FT DISULFID 200 244  
 FT DISULFID 230 257  
 FT DISULFID 262 306  
 FT DISULFID 292 319  
 FT DISULFID 324 368  
 FT DISULFID 354 381  
 FT DISULFID 386 430  
 FT DISULFID 416 443  
 FT DISULFID 448 492  
 FT DISULFID 478 505  
 FT DISULFID 510 554  
 FT DISULFID 540 567  
 FT DISULFID 580 624  
 FT DISULFID 610 637  
 FT DISULFID 642 686  
 FT DISULFID 672 699  
 FT CARBOHYD 398 398  
 FT CARBOHYD 603 603  
 FT CARBOHYD 654 654  
 FT CARBOHYD 661 661  
 FT CARBOHYD 679 679  
 FT LIPID 745 745  
 FT SITE 756 759  
 FT CONFLICT 724 724  
 FT SEQUENCE 768 AA; 83098 MW; B5173074D2F66B68 CRC64;

Query Match 2.2%; Score 219; DB 1; Length 768;  
 Best Local Similarity 23.1%; Pred. No. 6.1e-07;  
 Matches 99; Conservative 38; Mismatches 146; Indels 146; Gaps 29;  
 1384 SCIRPGKQKSCSLLLDHADVNCSTSIGGLKKKALTCRGFAIQAASSGQYIRPMQK-1442

Db 162 SCQDMGNSNGECIERI-----GSYTC--SCYPRF--YGPCEVYKEGKV 203  
 QY 1443 ----EILITCSS--GHMDONVSLPYDCGVPSLVNVAIFSCSEGTFLKRCISICVPP 1496  
 Db 204 NIPQVLMNCSHPLGERSFNSQC-----TFSCAEGVEL-----DGP 239  
 QY 1497 AKIQGLSPMLTCEDEGLMSPEVYC-KLEC---DAPPIIINALLPH-----CLQDNHDV 1548  
 Db 240 GELQ-----CLASGIMTNPPKCDVAVOCOSLEAP-----PHGTMACM--HPI 279  
 QY 1549 GTI-----CKEYCKPGYVVAESAEGKVRNKLKIQCLEGGIWEQ--GSCIPIVCEPPPV 1601  
 Db 280 AAFAYVSSCKFEQCPGYRAGSN-----TLHCTGSSQNSBPLPTCAIACEPPEIP 330  
 QY 1602 FEGWYEC--TNGPSLDSQCVLNCOE--REKLPILCTEGMTGWFKLCENLGSCP- 1654  
 Db 331 IHGSMDCVPSTGFYGVNSCTPLCAEGFVLKGNDAICADSGQWTAPAFCEALQ--CPE 388  
 QY 1655 -PPSELNSVRYKCEQCYG--IGAVCSPLCVIPPSDPVNLPEKITADTLEHMEPYKVQ 1710  
 Db 389 FPVPSK--AQVNCSPFGTLTYQSVCSFSC--DEGSIL-----VGAS 426  
 QY 1711 SIYCTGRORH--PPEVLVHCIOCEPFOADGWC-DTINNRAV---CHY--DGC----- 1756  
 Db 427 VIRCLATGHNNGAPRECAVSCAPMLSPENGSMTCVQPLGNSYKTCQPMCDGFFYLSG 486  
 QY 1757 ----DCCSS 1761  
 Db 487 PERLDCCPS 495  
 RESULT 15  
 CR2\_MOUSE STANDARD; PRT; 1025 AA.  
 ID CR2\_MOUSE  
 AC P19070;  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 01-NOV-1990 (Rel. 16, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Complement receptor type 2 precursor (Cr2) (Complement C3d receptor).  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OC NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BALB/c;  
 RX MEDLINE=90229735; PubMed=2139457;  
 RA Fingerboth J.D.;  
 RT "Comparative structure and evolution of murine CR2. The homolog of  
 the human C3d/BBV receptor (CD21).";  
 RL J. Immunol. 144:3458-3467(1990).  
 RP [2]  
 RP SEQUENCE OF 12-1025 FROM N.A.  
 RX MEDLINE=91010789; PubMed=2145366;  
 RA Molina H., Khoshnita T., Inoue K., Carel J.C., Holera V.M.;  
 RT "A molecular and immunochemical characterization of mouse CR2.  
 Evidence for a single gene model of mouse complement receptors 1 and  
 2.";  
 RL J. Immunol. 145:12974-2983(1990).  
 RP [3]  
 RP SEQUENCE OF 343-401 AND 991-1025 FROM N.A.  
 RX MEDLINE=89098890; PubMed=2783485;  
 RA Fingerboth J.D., Benedict M.A., Levy D.N., Strominger J.L.;  
 RT "Identification of murine complement receptor type 2.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 86:242-246(1989).  
 RP [4]  
 RP SEQUENCE OF 289-1025 FROM N.A.  
 RX MEDLINE=89381350; PubMed=2528587;  
 RA Kutz C.B., Paul M.S., Aegeerter M., Weis J.J., Weis J.H.;  
 RT "Murine complement receptor gene family. II. Identification and  
 characterization of the murine homolog (Cr2) to human CR2 and its

[illegible]

FT	DISULFID	709	752	BY SIMILARITY.
FT	DISULFID	738	769	BY SIMILARITY.
FT	DISULFID	778	820	BY SIMILARITY.
FT	DISULFID	806	833	BY SIMILARITY.
FT	DISULFID	841	884	BY SIMILARITY.
FT	DISULFID	870	897	BY SIMILARITY.
FT	DISULFID	902	945	BY SIMILARITY.
FT	DISULFID	931	958	BY SIMILARITY.
FT	CARBOHYD	77	77	N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	CARBOHYD	113	113	N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	CARBOHYD	276	276	N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	CARBOHYD	316	316	N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	CARBOHYD	364	364	N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	CARBOHYD	380	380	N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	CARBOHYD	484	484	N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	CARBOHYD	527	527	N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	CARBOHYD	615	615	N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	CARBOHYD	639	639	N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	CARBOHYD	694	694	N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	CARBOHYD	754	754	N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	CARBOHYD	790	790	N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	CARBOHYD	813	813	N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	CARBOHYD	823	823	N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	CARBOHYD	851	851	N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	CARBOHYD	901	901	N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	CONFLICT	289	291	YGS -> EFR (IN REF. 4).
FT	CONFLICT	306	306	S -> T (IN REF. 2).
FT	CONFLICT	520	520	P -> A (IN REF. 2).
FT	CONFLICT	962	963	MISSING (IN REF. 4).
QQ	SEQUENCE	1025 AA,	112994 MW,	19E518B9A0273694 CRC64;

Query Match	2.2%;	Score 218.5;	DB 1;	Length 1025;
Best Local Similarity	18.1%;	Pred. No. 1e-06;		
Matches	209;	Conservative 115;	Mismatches 389;	Indels 443; Gaps 57;
Qy	732	IHEVGHVGLYHV--FKGVSRSSCND---PCKETVPSMETGDLCADTAPTEKSELCR	784	
Db	152	IHN-GHHTG-OHVDQFVAGLSVTYSCBPGYLLTKKTIKICLSSD-MGVLPTECAOCE	208	
Qy	785	EP-----EPTSDCGFTFRPGAFPTNYMSYTDNCTDNFTPNQVARNHCYADLYOQ	835	
Db	209	HPGKFRPGQVKEPFLSLQVGTIV-----FSCNCGYOLQOQPSGQVIVVEKA	255	
Qy	836	QWTESRKPTPIPIPPWVIGQTKSLTTHMLPPISGVYVDRASSGLCGACTEDGTFRQYVH	895	
Db	256	IWTKKPVCKEILCPPP-----PPV-----RNGS-----HTGSRSENVPR	288	
Qy	896	TASRRVCDSSGYVTPBEAVGPDPDQCPCEPSLQAMSPVHLHYMNMTVPCEPTECSLEL	955	
Db	289	YGST-----VYTTCDSS-----PEKGVSTL	309	
Qy	956	LFOHPVQADTLTLMVTSFFMRESSGVLPDTEILLENKSSVHLGRLDPFCDIPLYTIKLVHDG	1015	
Db	310	IGETKINCTT-----GSGQ-----KTGIWGGPA-PLYVLSTSAVLCLQAP	346	
Qy	1016	KVSEVKV-----YTFDERIDEA-----ALTSOPHSPLC-SGCRPVR	1052	
Db	347	KIKRGQILLISIKDYSYNDIVASFCSEGFILKGRSTRCNAGHWBPVPVPCBKCO---	403	
Qy	1053	YQVLRDPPFASGLPVVYTHSHRKETDVEVTPGQWYQYQVLAEGAGELGASRPLNHHIGA	1112	
Db	404	-----APPKINGQKEDSYLNFDPGTSIRY-----	429	
Qy	1113	PYCGDGKVSRLGECDGDLVSDGSGSKVCELEEGFRCVGEBSPLCYVBEODGICEPPER	1172	
Db	430	-----SCDPGYLVG-----EDTILCTPE-----GKMTPTTP	456	
Qy	1173	KTSIVDC--GIYTPKGYLDQMATRAVSSHEDKKKCPVSLVTGSRPHSLICTSYHDLBNH	1222	
Db	457	QCTVAECKRPVPHLFXRPQNGFIRTAIVMS--SCDEGFGDLSAVGLCGGTLP----	506	
Qy	1230	RPLTGMF-----PCVASENETODDRSBQPEGSLKKEDEVMKLVCCNRPRGEA	1275	

Db 507 -----WFIRLCKEITCPEPPVHNGHTWSSSEDVPGT-----VVTYMCYGPBEG 555  
QY 1276 RAI-----FIFLTDLVPGEHQPTVTLYL-----TDVRGSNHSLGTYGLSCQH- 1321  
Db 556 VKFKLGEQTHCTSDSRGRGSSPAPLCKSLPANOCTDVHEN-----GVKLTDK 609  
QY 1322 -PLIINVTHQONVLFHHTTSVLNFPSSPRVGISAVALRTSSRIGLSAPNSCISEBQNH 1380  
Db 610 APFEYN-----DSVWFKCDGDGYILSSSQ-----IRCKANNMTWDE----- 645  
QY 1381 QGQSCIHRCGKODSCPSLL-----DHADVNCSTISGELMKCAITCQGFALQASSG 1434  
Db 646 -----KPLCKKEGCEPRMVRHGLPDSHTKLVK-----RTCONGYOL-----TG 683  
QY 1435 QYIRPMOKELITCQSSGHMDQNVSCLPVDCGVDPDSLNV-----YAN---FSC 1479  
Db 684 YTEKCO-----NAENGTWFKKIEVCTVILCQPPKIANGHTGMMAKFLYGENEVSIEC 738  
QY 1480 SEGTKFLKRCISICVPPAKLQGLSPWLTCLEDLMSLPEVYC-----KLECDAPPIILNA 1534  
Db 739 DEGFYLLGERKSLQCVNDSK-----GHGSMGPPQCLOQSPUTHCPDPEVAGHY 787  
QY 1535 NULLPHCLQDNHDVGTCKYECKRGYVVASABGKVRNKLKIQCLEGIMEQ--SCT- 1591  
Db 788 KLNKTHSAFSHND---IVHFVNOGF-----INMGSHLIRCHTNNTWLPDGPVTCIR 835  
QY 1592 --PVVCEPP-----PPVPGMYECTNGFSLDSQCVLNCNOERKLPILC 1633  
Db 836 KASIGCOSFSTIRNGNHTGSIARFPQMSVMSYQGFIMAGEARL-----IC 884  
QY 1634 TKEGLWTQEFKLCENLQGECP--PSELNSVEYKCEQ--YGIGAVCSPLCV 1682  
Db 885 THSGTWSQ-----PPFCKEVNCSFPEDTNGIQKGPQKTYRFGATVITLEC- 931  
QY 1683 IPPSDPVM---PENITADTLEHMEBPVKQSVICTGRQWHPDVL--VHCIOSECFQ 1737  
Db 932 ---EDGYTLEGSPQSCODD--SQWNPPL---ALCKYR--WSTIPLIGISVGSALITIM 982  
QY 1738 ADGWCDTINRAYCHY 1753  
Db 983 SVGFCMILKRESNYY 998

Search completed: January 2, 2004, 16:03:18  
Job time : 27 secs

**THIS PAGE BLANK (USPTO)**



FT Binding-site 733..743  
 /note= "Catalytic zinc binding motif"  
 FT Binding-site 805..809  
 /note= "Catalytic zinc binding motif"  
 FT Region 1396..1459  
 /note= "Short consensus repeat (SCR-1)"  
 FT Region 1464..1521  
 /note= "Short consensus repeat (SCR-2)"  
 FT Region 1525..1590  
 /note= "Short consensus repeat (SCR-3)"  
 FT Region 1595..1646  
 /note= "Short consensus repeat (SCR-4)"  
 FT Region 1653..1729  
 /note= "Short consensus repeat (SCR-5)"  
 FT Domain 1733..1758  
 /note= "Lin- notch motif (LNR3)"  
 XX MO200232953-A2.  
 XX 25-APR-2002.  
 XX 19-OCT-2001; 2001WO-DK00695.  
 XX 20-OCT-2000; 2000DK-0001571.  
 XX 20-OCT-2000; 2000US-241840P.  
 PA (COMO-) COMO BIOTECH APS.  
 PI Oxvig C, Overgaard MT;  
 XX WPI; 2002-444239/47.  
 DR N-PSDB; AAD38203.  
 XX New pregnancy-associated plasma protein A2 and encoding nucleic acids,  
 PT useful as marker for pathological states (e.g. Down's syndrome), or as  
 PT a therapeutic target for drugs that modify the proteolytic activity of  
 PT PAP-A2  
 XX Claim 12; Fig 1; 113pp; English.  
 XX The invention relates to pregnancy-associated plasma protein A2 (PAP-A2)  
 CC and its corresponding nucleic acid. PAP-A2 is useful as a marker  
 CC for pathological states (e.g. Down's syndrome, acute coronary syndrome,  
 CC unstable angina or myocardial infarction), and is used to treat diseases  
 CC such as restenosis, coronary atherosclerosis, osteoporosis and cancer.  
 CC It is used as a therapeutic target for drugs that modify the proteolytic  
 CC activity of PAP-A2 in pregnant and non-pregnant individuals and it is  
 CC also used in wound healing. The PAP-A2 cDNA is useful for producing  
 CC full length PAP-A2, and the PAP-A2 gene can be used for gene therapy  
 CC to introduce PAP-A2 into target cells. The PAP-A2 proteins can be used  
 CC in the production of antibodies, diagnostic kits, laboratory reagents and  
 CC assays. The identification of compounds that affect PAP-A2 function;  
 CC as a therapeutic target for the reduction or elimination of IGFBP-5  
 CC proteolytic activity in a cell; and as antigen for immunization. The  
 CC inhibitory agent or enhancing agent is useful in the manufacture of a  
 CC medicament for treating a clinical condition in an individual in need  
 CC of such treatment. The present sequence is human preproPAP-A2 protein.  
 CC PAP-A2 is an active enzyme which cleaves IGFBP-5.  
 XX  
 SQ Sequence 1791 AA;  
 Query Match 100.0%; Score 1791; DB 23; Length 1791;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1791; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 AVEBPAPWVGDSPIGSELLGDDDAVLGNQSKESLGEAGIIOKSGMAAATTTAIFTTL 180  
 DB 121 AVEBPAPWVGDSPIGSELLGDDDAVLGNQSKESLGEAGIIOKSGMAAATTTAIFTTL 180  
 QY 181 NEPKETQRRGMWAKSRQRRQWKRARADGGQDSGISSHFQWMPKHSIKHYKXSPPEESN 240  
 DB 181 NEPKETQRRGMWAKSRQRRQWKRARADGGQDSGISSHFQWMPKHSIKHYKXSPPEESN 240  
 QY 241 QNGEGSYRBAETNSQVGPILYFSGRRELLLRPVLAIEIPBAFTVEAMVPEEGON 300  
 DB 241 QNGEGSYRBAETNSQVGPILYFSGRRELLLRPVLAIEIPBAFTVEAMVPEEGON 300  
 QY 301 NPALIAVFDNCSHTVSDKGWALGIRSGKXGRDAEPFSLCTDRYKAKATILISHSRQ 360  
 DB 301 NPALIAVFDNCSHTVSDKGWALGIRSGKXGRDAEPFSLCTDRYKAKATILISHSRQ 360  
 QY 361 PGTWTHAATYDGNHMLLYDGTQVASSLDQSGPLNSPPMASSCSLLDGDSSDGHYFR 420  
 DB 361 PGTWTHAATYDGNHMLLYDGTQVASSLDQSGPLNSPPMASSCSLLDGDSSDGHYFR 420  
 QY 421 GHLGTIVFWSSTALPQSHRQSSQSSGSEBEATDVLVLAEPVNTWVPPRDEKYPRLV 480  
 DB 421 GHLGTIVFWSSTALPQSHRQSSQSSGSEBEATDVLVLAEPVNTWVPPRDEKYPRLV 480  
 QY 481 LQGFEPPEILSPLOPPLCGQTVCDNVELISQYNGWPLRGEKVIYQVNNICDEGLNP 540  
 DB 481 LQGFEPPEILSPLOPPLCGQTVCDNVELISQYNGWPLRGEKVIYQVNNICDEGLNP 540  
 QY 541 IVSEBQIRLOHEALNEAFSRYNISWQLSVHOVNSTLRHRVTVLNCBPSKIGNDHCP 600  
 DB 541 IVSEBQIRLOHEALNEAFSRYNISWQLSVHOVNSTLRHRVTVLNCBPSKIGNDHCP 600  
 QY 541 IVSEBQIRLOHEALNEAFSRYNISWQLSVHOVNSTLRHRVTVLNCBPSKIGNDHCP 600  
 DB 541 IVSEBQIRLOHEALNEAFSRYNISWQLSVHOVNSTLRHRVTVLNCBPSKIGNDHCP 600  
 QY 601 EHPITGVGDGCRLOGCYSWNRBDGCHVECNMMLNDFDGDCCDQVADVARTCTCPDP 660  
 DB 601 EHPITGVGDGCRLOGCYSWNRBDGCHVECNMMLNDFDGDCCDQVADVARTCTCPDP 660  
 QY 601 EHPITGVGDGCRLOGCYSWNRBDGCHVECNMMLNDFDGDCCDQVADVARTCTCPDP 660  
 DB 601 EHPITGVGDGCRLOGCYSWNRBDGCHVECNMMLNDFDGDCCDQVADVARTCTCPDP 660  
 QY 661 SPKRAVSVKELKALQNLNSTHPLNTYFASVREDLAGAATPMDKAVTHLGGIVLSPA 720  
 DB 661 SPKRAVSVKELKALQNLNSTHPLNTYFASVREDLAGAATPMDKAVTHLGGIVLSPA 720  
 QY 721 YYGMPGHTDTHIEVGHVGLIYHFKGVSEBESCNDECKETVBSMEFGDLCADTAIPPKS 780  
 DB 721 YYGMPGHTDTHIEVGHVGLIYHFKGVSEBESCNDECKETVBSMEFGDLCADTAIPPKS 780  
 QY 781 ELCEBPEPTSTGCTFPFGAPFTNMSYTDNCTDNFTPQVARMHCYLDLVYQOMTES 840  
 DB 781 ELCEBPEPTSTGCTFPFGAPFTNMSYTDNCTDNFTPQVARMHCYLDLVYQOMTES 840  
 QY 841 RKPTPIPIPPVVIQTNKSLTIHMLPISGVVYDRASGSLGACTBDGTFRQYVHTASSR 900  
 DB 841 RKPTPIPIPPVVIQTNKSLTIHMLPISGVVYDRASGSLGACTBDGTFRQYVHTASSR 900  
 QY 901 RVCDSSGWTPEEAVGPPVDVQCEPSIQANSPEVHLIYHMMYVPCPTBEGCSLELFOHP 960  
 DB 901 RVCDSSGWTPEEAVGPPVDVQCEPSIQANSPEVHLIYHMMYVPCPTBEGCSLELFOHP 960  
 QY 961 VQADTLTMTWTSFPMESQVLFDEILLENESVHLGPDTFDCDPLITIKLHVQKXSGV 1020  
 DB 961 VQADTLTMTWTSFPMESQVLFDEILLENESVHLGPDTFDCDPLITIKLHVQKXSGV 1020  
 QY 1021 KYTFDERIEIDALLTSQPHSPICSGCPVRYOVULNDPPASGLPVVYVTHSHKFTDVE 1080  
 DB 1021 KYTFDERIEIDALLTSQPHSPICSGCPVRYOVULNDPPASGLPVVYVTHSHKFTDVE 1080  
 QY 1081 VTPGQMYOVYLAAGELGASPLNHIHAPYCGDGKYSERLAGEBCDDGLVSGDGS 1140  
 DB 1081 VTPGQMYOVYLAAGELGASPLNHIHAPYCGDGKYSERLAGEBCDDGLVSGDGS 1140  
 QY 1141 KVCELBEFGNVCBPSLCYMEBGDGCIPFERKTSYVDCGIYTKRGYLDQATAYSSHE 1200  
 DB 1141 KVCELBEFGNVCBPSLCYMEBGDGCIPFERKTSYVDCGIYTKRGYLDQATAYSSHE 1200  
 QY 1201 DKKCPVSLVYGBPHSLICTSYHNDLPVHRPLTGMFPCVASENETQDDBSBOPEGLAKE 1260



```

Db      1201 DKKKCPVSLVTGEBHSICTSYHPDLTPNHRPLTGMFPCVASENETQDRSEQEGSLKKE 1260
Qy      1261 DEWMLKVCENRPGBARLIFELTTDGLVPEHOOPTVTLVTVRGSNHSIGTYGLSCOH 1320
Db      1261 DEWMLKVCENRPGBARLIFELTTDGLVPEHOOPTVTLVTVRGSNHSIGTYGLSCOH 1320
Qy      1321 NPLIINTVHHQNVLFHHTTSTVLNFSPPRGVISAVALRTSSRIGLSAPNSCISEDEQNH 1380
Db      1321 NPLIINTVHHQNVLFHHTTSTVLNFSPPRGVISAVALRTSSRIGLSAPNSCISEDEQNH 1380
Qy      1381 QGQSCIHRCGKQDSCPSLLLDHADVNTCTSIGBLMKCAITCGRGALQASSGQYIRPM 1440
Db      1381 QGQSCIHRCGKQDSCPSLLLDHADVNTCTSIGBLMKCAITCGRGALQASSGQYIRPM 1440
Qy      1441 QKEILLTCSGSHMDQNVSCLPVDCGVPDPSLVNANFSCBGTGFKRCSISCVPAKLQ 1500
Db      1441 QKEILLTCSGSHMDQNVSCLPVDCGVPDPSLVNANFSCBGTGFKRCSISCVPAKLQ 1500
Qy      1501 GLSPMLTCLBGLMSLEPVYCKLECDAPITILNANLLPHCLQDNHVDGTCIKYCKRPGY 1560
Db      1501 GLSPMLTCLBGLMSLEPVYCKLECDAPITILNANLLPHCLQDNHVDGTCIKYCKRPGY 1560
Qy      1561 YVABSAEGKVRNKLKIQCLEGGIWEQSCIPVCEBPVFEQMYECTNGFSLDSCVL 1620
Db      1561 YVABSAEGKVRNKLKIQCLEGGIWEQSCIPVCEBPVFEQMYECTNGFSLDSCVL 1620
Qy      1621 NCOERKLPILCTKEBLMTQEFCLCNLQEGCEPPSELSNVSKYKQGYIGAVCSPL 1680
Db      1621 NCOERKLPILCTKEBLMTQEFCLCNLQEGCEPPSELSNVSKYKQGYIGAVCSPL 1680
Qy      1681 CVIPSPDPMVLPENITADTLEHMMEPVKVOSIVCTGRQHPDVLVHCIOGCEPQADG 1740
Db      1681 CVIPSPDPMVLPENITADTLEHMMEPVKVOSIVCTGRQHPDVLVHCIOGCEPQADG 1740
Qy      1741 WCDTINNRAYCHVDGDCSSSTLSKTVIPPAADCDDECTCRDPAEENQ 1791
Db      1741 WCDTINNRAYCHVDGDCSSSTLSKTVIPPAADCDDECTCRDPAEENQ 1791

RESULT 2
AAE23826
ID AAE23826 standard; Protein; 1791 AA.
XX
AC AAE23826;
XX
DT 10-SEP-2002 (first entry)
XX
DE Human preproPAP-A2 protein mutant, E734Q.
XX
KW Human; pregnancy-associated plasma protein A2; PAP-A2; Down's syndrome;
KW acute coronary syndrome; angina; myocardial infarction; restenosis;
KW coronary atherosclerosis; wound healing; osteoporosis; cancer; cadiant;
KW gene therapy; IGFBP-5; immunisation; gynaecological; antiinflammatory;
KW vasotrophic; nocrotropic; cytoskeletal; antiarteriosclerotic; vulneryary;
KW osteopathic; enzyme; mutant; mutein.
XX
OS Homo sapiens.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Region 1..233
FT Peptide /note= "Human PAP-A2 prepro region"
FT Region 23..233
FT Region /label= Signal_peptide
FT Region 1..22
FT Region /note= "Human PAP-A2 pro region"
FT Region 23..1791
FT Protein /note= "Human mature PAP-A2 protein"
FT Domain 586..612
FT Domain /note= "Lin-notch motif (LNR1)"
FT Domain 618..644
FT Domain /note= "Lin-notch motif (LNR2)"

```

```

FT Binding-site 733..743
FT /note= "Catalytic zinc binding motif"
FT Misc-difference 734
FT /note= "Wild-type Glu substituted with Gln"
FT Binding-site 805..809
FT /note= "Catalytic zinc binding motif"
FT Region 1396..1459
FT /note= "short consensus repeat (SCR-1)"
FT Region 1464..1521
FT /note= "short consensus repeat (SCR-2)"
FT Region 1525..1590
FT /note= "short consensus repeat (SCR-3)"
FT Region 1595..1646
FT /note= "short consensus repeat (SCR-4)"
FT Region 1653..1729
FT /note= "short consensus repeat (SCR-5)"
FT Domain 1733..1758
FT /note= "Lin-notch motif (LNR3)"
XX
PN WO200232953-A2.
XX
PD 25-APR-2002.
XX
XX 19-OCT-2001; 2001WO-DK00695.
XX
XX 20-OCT-2000; 2000DK-0001571.
XX 20-OCT-2000; 2000US-241840P.
XX
XX (COMO-) COMO BIOTECH APS.
XX
XX Oxvig C, Overgaard MT;
XX
XX WPI: 2002-444239/47.
XX DR N-PSDB; AAD38324.
XX
XX New pregnancy-associated plasma protein A2 and encoding nucleic acids,
XX useful as marker for pathological states (e.g. Down's syndrome), or as
XX a therapeutic target for drugs that modify the proteolytic activity of
XX PAP-A2.
XX
XX Example ; Page -; 113pp; English.
XX
XX The invention relates to pregnancy-associated plasma protein A2 (PAP-A2)
XX and its corresponding nucleic acid. PAP-A2 is useful as a marker
XX for pathological states (e.g. Down's syndrome, acute coronary syndrome,
XX unstable angina or myocardial infarction), and is used to treat diseases
XX such as restenosis, coronary atherosclerosis, osteoporosis and cancer.
XX It is used as a therapeutic target for drugs that modify the proteolytic
XX activity of PAP-A2 in pregnant and non-pregnant individuals and it is
XX also used in wound healing. The PAP-A2 cDNA is useful for producing
XX full length PAP-A2, and the PAP-A2 gene can be used for gene therapy
XX to introduce PAP-A2 into target cells. The PAP-A2 proteins can be used
XX in the production of antibodies, diagnostic kits, laboratory reagents and
XX assays; in the identification of compounds that affect PAP-A2 function;
XX as a therapeutic target for the reduction or elimination of IGFBP-5
XX proteolytic activity in a cell; and as antigen for immunization. The
XX inhibitory agent or enhancing agent is useful in the manufacture of a
XX medicament for treating a clinical condition in an individual in need
XX of such treatment. The present sequence is human preproPAP-A2 protein
XX mutant. PAP-A2 is an active enzyme which cleaves IGFBP-5.
XX Note: This sequence is not shown in the specification, however it
XX is constructed based on human preproPAP-A2 protein shown in fig 1
XX (AAE23821).
XX
SQ Sequence 1791 AA;
XX
Query Match 94.4%; Score 1690; DB 23; Length 1791;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1790; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 MMCKILRISLALAGMALCSANSELGWTRKSLVEREHLNQVLGEGRCWLGAKYRRPR 60
Db 1 MMCKILRISLALAGMALCSANSELGWTRKSLVEREHLNQVLGEGRCWLGAKYRRPR 60

```

```

Qy 61 ASPQHLLFGVYPSRAGNYLRPYVGEQEIHTHGRSKEDTEGNAVSLVPEPDLTENPAGLRG 120
Db 61 ASPQHLLFGVYPSRAGNYLRPYVGEQEIHTHGRSKEDTEGNAVSLVPEPDLTENPAGLRG 120
Qy 121 AVEBPAAWVGDSPITGSEELIGDDDAYIGNORSKESLGEAGIQKGSAMAATTTTAIFTTL 180
Db 121 AVEBPAAWVGDSPITGSEELIGDDDAYIGNORSKESLGEAGIQKGSAMAATTTTAIFTTL 180
Qy 181 NEPEPQRGRMASSRORROYMKRRAEDQGDGSGISHPQWPKHSIKHRVKSPPRESN 240
Db 181 NEPEPQRGRMASSRORROYMKRRAEDQGDGSGISHPQWPKHSIKHRVKSPPRESN 240
Qy 241 QNGEGSYREAEFTENSQVGLPILYFSGRERELLREPEVLAIEIPREAFTEAMVAREGON 300
Db 241 QNGEGSYREAEFTENSQVGLPILYFSGRERELLREPEVLAIEIPREAFTEAMVAREGON 300
Qy 301 NPALIAGVFNDCSTVSDKGMALGIRSGKDKGKDARFPFSLCTDRYKATILISHSRQ 360
Db 301 NPALIAGVFNDCSTVSDKGMALGIRSGKDKGKDARFPFSLCTDRYKATILISHSRQ 360
Qy 361 PGTTHVAATYDGHMMLYDGTQVASSLDQSGPLNSPFMASCSLLIGDSSSDGHYR 420
Db 361 PGTTHVAATYDGHMMLYDGTQVASSLDQSGPLNSPFMASCSLLIGDSSSDGHYR 420
Qy 421 GHGLTLVFWSTALPQSHFOHSSQHSGBEATDLVLTASFEPVNTWVPPFDEKYPLEV 480
Db 421 GHGLTLVFWSTALPQSHFOHSSQHSGBEATDLVLTASFEPVNTWVPPFDEKYPLEV 480
Qy 481 LGGEPEPEIISPIQPLCGQTCVNDNELISQYNGWPLRGEKVIYQVANI CDEGLNP 540
Db 481 LGGEPEPEIISPIQPLCGQTCVNDNELISQYNGWPLRGEKVIYQVANI CDEGLNP 540
Qy 541 IVSEEOIRLOHEALNEAFSRNISMOLSVHVNSTLPHRVLYNCEBSKIGNHCPBC 600
Db 541 IVSEEOIRLOHEALNEAFSRNISMOLSVHVNSTLPHRVLYNCEBSKIGNHCPBC 600
Qy 601 EHPPLTGYDGDGCRLOGRCSYNNRRDGLCHVECNMMLNFDGDCDDPOVAVRKTCEPD 660
Db 601 EHPPLTGYDGDGCRLOGRCSYNNRRDGLCHVECNMMLNFDGDCDDPOVAVRKTCEPD 660
Qy 661 SPKRAYNSVKELKALQLNSTHPLNIYFASVREDLAGAATPMDKQAVTHLGSI VLSPA 720
Db 661 SPKRAYNSVKELKALQLNSTHPLNIYFASVREDLAGAATPMDKQAVTHLGSI VLSPA 720
Qy 721 YXGMPGHTDMIHVGHVGLYHVFYKVSERESCNDCPKETVPBMETGDLCAADTAPTPKS 780
Db 721 YXGMPGHTDMIHVGHVGLYHVFYKVSERESCNDCPKETVPBMETGDLCAADTAPTPKS 780
Qy 781 ELCEPEPSTDTGCFTRPGAPFTNYSYTDNCTDNFTPNQVARMHCYLDLVYQOMTES 840
Db 781 ELCEPEPSTDTGCFTRPGAPFTNYSYTDNCTDNFTPNQVARMHCYLDLVYQOMTES 840
Qy 841 RKPTPIPIPPNVVIGQTNKSLTIHMLPISGVVYRASSGLCGACTEDGTFRQYVHTASSR 900
Db 841 RKPTPIPIPPNVVIGQTNKSLTIHMLPISGVVYRASSGLCGACTEDGTFRQYVHTASSR 900
Qy 901 RVCSGSGVWPEBAVGPDPVQPCPSLOAMSPEVHLVHMMVTPCPTEGSLBLFQHP 960
Db 901 RVCSGSGVWPEBAVGPDPVQPCPSLOAMSPEVHLVHMMVTPCPTEGSLBLFQHP 960
Qy 961 VOADTLTLMVTSFMESSQVLPDTEILLENKESVHGLPDTFCDIPLTIKLVHVGKSGV 1020
Db 961 VOADTLTLMVTSFMESSQVLPDTEILLENKESVHGLPDTFCDIPLTIKLVHVGKSGV 1020
Qy 1021 KYVTFDEIEIDALLTSQPSPLCSGCRPVRYQVLADPPASGLPVVVTSHSKFTDVE 1080
Db 1021 KYVTFDEIEIDALLTSQPSPLCSGCRPVRYQVLADPPASGLPVVVTSHSKFTDVE 1080
Qy 1081 VTPGOMYQVYLAEGELGASPLNHIHGAIPYCGDGKVBRLGEBECDDGLVSGDCS 1140
Db 1081 VTPGOMYQVYLAEGELGASPLNHIHGAIPYCGDGKVBRLGEBECDDGLVSGDCS 1140

```

```

Qy 1141 KCVELEBGFNCVGEPSLCYMEBGDI CEPEBKRTSI VDCGIYTPKGYLDQMATRAYSSHE 1200
Db 1141 KCVELEBGFNCVGEPSLCYMEBGDI CEPEBKRTSI VDCGIYTPKGYLDQMATRAYSSHE 1200
Qy 1201 DKKKCPVSLVTGPHSLICTSYHPDLPNRPLTGMFPVCASENETDDDSBPQEGSLKKE 1260
Db 1201 DKKKCPVSLVTGPHSLICTSYHPDLPNRPLTGMFPVCASENETDDDSBPQEGSLKKE 1260
Qy 1261 DEYMLKVCENRPEBARAIFILFTTDLVPEGHQOPVTIYLTDPVRSNLSLTYGSLCOH 1320
Db 1261 DEYMLKVCENRPEBARAIFILFTTDLVPEGHQOPVTIYLTDPVRSNLSLTYGSLCOH 1320
Qy 1321 NPLIINTVTHQNVLPHTTSSVLLNFSSPRVGSAAVALRTSSRIGLSAPSNCISEDEGQNH 1380
Db 1321 NPLIINTVTHQNVLPHTTSSVLLNFSSPRVGSAAVALRTSSRIGLSAPSNCISEDEGQNH 1380
Qy 1381 QGOSCIHRPCGKODSCPSLLDHDADVNCISIGPGLMKCAITCQRPALQASSGOYIRPM 1440
Db 1381 QGOSCIHRPCGKODSCPSLLDHDADVNCISIGPGLMKCAITCQRPALQASSGOYIRPM 1440
Qy 1441 QKEILLTSSSGHMDQNVSCLPVDCGVPDPSLVNYANFSCSEGTKFLKCSISGVPRAKLQ 1500
Db 1441 QKEILLTSSSGHMDQNVSCLPVDCGVPDPSLVNYANFSCSEGTKFLKCSISGVPRAKLQ 1500
Qy 1501 GLSPMLTCLBEGDGLMSLPEVYCKLECDAPRIILNANLLPHCLQDNHDVGTICKYECKPGY 1560
Db 1501 GLSPMLTCLBEGDGLMSLPEVYCKLECDAPRIILNANLLPHCLQDNHDVGTICKYECKPGY 1560
Qy 1561 YVABSAAGKRNKLLKIQCLEGGIWEQSGCIPVCGPPPPVFGMTECTNGBSLDQCVL 1620
Db 1561 YVABSAAGKRNKLLKIQCLEGGIWEQSGCIPVCGPPPPVFGMTECTNGBSLDQCVL 1620
Qy 1621 NQNOERKPLILCTKGMLTQEFKLCBNLQGECPRPPELSNVEYKCEQGYIGAVCSPL 1680
Db 1621 NQNOERKPLILCTKGMLTQEFKLCBNLQGECPRPPELSNVEYKCEQGYIGAVCSPL 1680
Qy 1681 CVIPSPDPVMLPENITADTLHEHMEBPVKQSIYCTGRQWHPDVLVHCIOGCEPQADG 1740
Db 1681 CVIPSPDPVMLPENITADTLHEHMEBPVKQSIYCTGRQWHPDVLVHCIOGCEPQADG 1740
Qy 1741 WCOTINRAYCHIDGDCSSTLSSKKVTPPADCDLDECTCRDPAABENQ 1791
Db 1741 WCOTINRAYCHIDGDCSSTLSSKKVTPPADCDLDECTCRDPAABENQ 1791

RESULT 3
ABB84488
ID ABB84488 standard; Protein; 1791 AA.
XX AC ABB84488;
XX DT 24-DEC-2002 (first entry)
XX DE Human pregnancy associated protein PAP-8a SEQ ID 3.
XX KW PAPP-B; human; pregnancy associated plasma protein E; abortive;
XX KW contraceptive; gene therapy; vaccine; pregnancy; antenatal; diagnosis;
XX KW dyogenetic pregnancy.
XX OS Homo sapiens.
XX PN US2002102252-A1.
XX PD 01-AUG-2002.
XX XX 06-APR-2001; 2001US-0827998.
XX XX 26-MAY-2000; 2000US-207456P.
XX XX (GUY/) GU Y.
XX XX (SHAN/) SHANNON M E.
XX XX Gu Y, Shannon ME;

```

XX WPI; 2002-697817/75.  
DR N-PSDB; ABS74483, ABS74484.  
XX  
PT New isolated nucleic acid encoding an isoform of human pregnancy  
associated plasma protein E, for preventing or aborting pregnancy  
XX  
PS Claim 1; Figure 3; 353pp; English.  
XX  
CC This invention describes a novel isolated nucleic acid that encodes  
CC one of three new isoforms of human pregnancy associated plasma protein E,  
CC hPAP-E. The products of the invention have abortive and contraceptive  
CC activity and can be used for gene therapy or in a vaccine. The nucleic  
CC acid, polypeptide encoded by it, or antibody to the polypeptide can be  
CC used in pharmaceutical compositions or vaccines for preventing or  
CC aborting pregnancy. PAP-E is used in the antenatal diagnosis of  
CC dysgenetic pregnancies. The nucleic acids are used as probes to assess  
CC the level of PAP-E isoform mRNA in chorionic villus samples, and the  
CC antibodies can be used to assess the expression levels of PAP-E isoform  
CC proteins in chorionic villus samples, to diagnose dysgenetic pregnancies  
CC antenatally. This sequence represents a human PAP-E protein described in  
CC the disclosure of the invention.

XX Sequence 1791 AA:

Query Match 85.8%; Score 1536; DB 23; Length 1791;

Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1736; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MMCKILIRISLAILAGALCSANSELGWTTRKSLVEREHLNOVLLEGERCMLGAKVRPR 60  
DB 1 MMCKILIRISLAILAGALCSANSELGWTTRKSLVEREHLNOVLLEGERCMLGAKVRPR 60  
QY 61 ASPOHHLFGVYPSAGVYLRPPYGEQIHHHTGSKPDTEGNVSLVPPDLTEPDLARG 120  
DB 61 ASPOHHLFGVYPSAGVYLRPPYGEQIHHHTGSKPDTEGNVSLVPPDLTEPDLARG 120  
QY 121 AVEBPAPWVGDSPIGQSELDGDDAYLGNQSKESIGEAGIQKSGMAATTTTATTTL 180  
DB 121 AVEBPAPWVGDSPIGQSELDGDDAYLGNQSKESIGEAGIQKSGMAATTTTATTTL 180  
QY 181 NEKPETORGMMAKSRORRYWKRRAADQSGSISHPQWPKSLKHKRVKSPRESN 240  
DB 181 NEKPETORGMMAKSRORRYWKRRAADQSGSISHPQWPKSLKHKRVKSPRESN 240  
QY 241 QNGGEGYRRAETNSQVGPILYFSGRERLLRPEVLAIRPEATTVEAKVPEEGON 300  
DB 241 QNGGEGYRRAETNSQVGPILYFSGRERLLRPEVLAIRPEATTVEAKVPEEGON 300  
QY 241 QNGGEGYRRAETNSQVGPILYFSGRERLLRPEVLAIRPEATTVEAKVPEEGON 300  
DB 241 QNGGEGYRRAETNSQVGPILYFSGRERLLRPEVLAIRPEATTVEAKVPEEGON 300  
QY 301 NPAILAGVFNCSHTVADKGMALGIRSGDKGRDARFFSLCTDRYKXATILISHSRQ 360  
DB 301 NPAILAGVFNCSHTVADKGMALGIRSGDKGRDARFFSLCTDRYKXATILISHSRQ 360  
QY 361 PGTWTHAATYDGRHMLLVYDGTQVASSLDQSGPLNSPFMASSCLLLGGDSSEGHYFR 420  
DB 361 PGTWTHAATYDGRHMLLVYDGTQVASSLDQSGPLNSPFMASSCLLLGGDSSEGHYFR 420  
QY 421 GHIGTLVFNSTALPQSHFQHSQHSQSGEATDLVLTASFEVNTVEVPRDEKYPLEV 480  
DB 421 GHIGTLVFNSTALPQSHFQHSQHSQSGEATDLVLTASFEVNTVEVPRDEKYPLEV 480  
QY 481 LOGFEPEPILSPLOPLCGQTVCDNVELISQVNGYMPRLGEXKIRQOVNICDDELNP 540  
DB 481 LOGFEPEPILSPLOPLCGQTVCDNVELISQVNGYMPRLGEXKIRQOVNICDDELNP 540  
QY 541 IYSEBOQIRLOHEALNEAFSRYNISWQSVHGVNSTLRHRVVLNCEPSKIGNDHCDPEC 600  
DB 541 IYSEBOQIRLOHEALNEAFSRYNISWQSVHGVNSTLRHRVVLNCEPSKIGNDHCDPEC 600  
QY 601 EHPILTYGDSGDCRLQSGCYSWNRDGLCHVECNMNLDFDGDCCDQVADVKTCTCPD 660  
DB 601 EHPILTYGDSGDCRLQSGCYSWNRDGLCHVECNMNLDFDGDCCDQVADVKTCTCPD 660

QY 661 SPKRAYSVKELKEALQNSTHFLNLYFASVREDLAAGATWMDKDAVTHLGIVLSPA 720  
DB 661 SPKRAYSVKELKEALQNSTHFLNLYFASVREDLAAGATWMDKDAVTHLGIVLSPA 720  
QY 721 YGMPGHTDMHEVGHVGLYHVFQVSEERECNDPCKEKTIVSMETGDLCAOTAPTPKS 780  
DB 721 YGMPGHTDMHEVGHVGLYHVFQVSEERECNDPCKEKTIVSMETGDLCAOTAPTPKS 780  
QY 781 ELCREPEPTSDTCGFRFPAGPPTNYSYTDNCTNFTDNOVABMHCYLDLYVQWMTES 840  
DB 781 ELCREPEPTSDTCGFRFPAGPPTNYSYTDNCTNFTDNOVABMHCYLDLYVQWMTES 840  
QY 841 RKPTPIPIPMVIGQTNKSLTTHMLPISGVVYDRASGSLCAGCTEDGTROYVHTASSR 900  
DB 841 RKPTPIPIPMVIGQTNKSLTTHMLPISGVVYDRASGSLCAGCTEDGTROYVHTASSR 900  
QY 901 RVCDSGYTPEBAVGPVVDQCEPSLQMSPEVHLYHNMVTPCCTECCSLELFOHP 960  
DB 901 RVCDSGYTPEBAVGPVVDQCEPSLQMSPEVHLYHNMVTPCCTECCSLELFOHP 960  
QY 961 VOADTLTLWTSFPMSSQVLPFTLELLENKESVHLGPIDTPCDIPLTLKLVHDKVSGV 1020  
DB 961 VOADTLTLWTSFPMSSQVLPFTLELLENKESVHLGPIDTPCDIPLTLKLVHDKVSGV 1020  
QY 1021 KYTTFDERIEIDALITSOHPSPLCSGCRPVRYQVLRDPPASGLPVVYTHSHRKFTDVE 1080  
DB 1021 KYTTFDERIEIDALITSOHPSPLCSGCRPVRYQVLRDPPASGLPVVYTHSHRKFTDVE 1080  
QY 1081 VTFQWYQVVLAEAGELGASPLNTHIGAPYCGDKVSERLGEECDGDLVSGDGS 1140  
DB 1081 VTFQWYQVVLAEAGELGASPLNTHIGAPYCGDKVSERLGEECDGDLVSGDGS 1140  
QY 1141 KVELEBEGFNCVEPSLCVWYEDGICEPERRTSLVDGCIYTPKGLDMARAYSHSHE 1200  
DB 1141 KVELEBEGFNCVEPSLCVWYEDGICEPERRTSLVDGCIYTPKGLDMARAYSHSHE 1200  
QY 1201 DKKKCPVSLVTGEPHSLICTSYHPDLPHNRPLTGMFPCVASENETODDRSEOPGSLKE 1260  
DB 1201 DKKKCPVSLVTGEPHSLICTSYHPDLPHNRPLTGMFPCVASENETODDRSEOPGSLKE 1260  
QY 1261 DEWVLKVCENRPEBADAIFLTITDGLVGEHOQPTVTLVTVRGSNHSLGTYGLSCQH 1320  
DB 1261 DEWVLKVCENRPEBADAIFLTITDGLVGEHOQPTVTLVTVRGSNHSLGTYGLSCQH 1320  
QY 1321 NPLIINTVHQNVLPHHTTSVNLNFSPPRGISAVNLRTSRIGLASPNSCISEDEQNH 1380  
DB 1321 NPLIINTVHQNVLPHHTTSVNLNFSPPRGISAVNLRTSRIGLASPNSCISEDEQNH 1380  
QY 1381 QGQSCIHRCPCGKODSCPSLLLDHADVYNCTSIGPLMKCAITCORGFALQASSGQYIRPM 1440  
DB 1381 QGQSCIHRCPCGKODSCPSLLLDHADVYNCTSIGPLMKCAITCORGFALQASSGQYIRPM 1440  
QY 1441 QKEILITCSGSHWDQVNSCLPVDGVPDPSLVNANFSCSEGTKFLKRCISICVPPAKQ 1500  
DB 1441 QKEILITCSGSHWDQVNSCLPVDGVPDPSLVNANFSCSEGTKFLKRCISICVPPAKQ 1500  
QY 1501 GLSPWILTCLEDDGMSLPBYVCKLECDAPRIILNANLLPHCLDNDHVGITCYKECPRGY 1560  
DB 1501 GLSPWILTCLEDDGMSLPBYVCKLECDAPRIILNANLLPHCLDNDHVGITCYKECPRGY 1560  
QY 1561 YVABSAAGVKNLTLKIOCLBEGIMEQSGCIPVCEBPVPVFGMEVCTNGFSLDSQCVL 1620  
DB 1561 YVABSAAGVKNLTLKIOCLBEGIMEQSGCIPVCEBPVPVFGMEVCTNGFSLDSQCVL 1620  
QY 1621 NCNQBEBKLPILCTKGLWTOEFKLCENLQEGCPRPPELSINVSVEYKCEQGYIGAVCSPL 1680  
DB 1621 NCNQBEBKLPILCTKGLWTOEFKLCENLQEGCPRPPELSINVSVEYKCEQGYIGAVCSPL 1680  
QY 1681 CVTPSPDVPWLPRNITADTLLEHMBEYKQVSIYCTGRQOMHPRPVVHCTIQSCPEPQA 1738  
DB 1681 CVTPSPDVPWLPRNITADTLLEHMBEYKQVSIYCTGRQOMHPRPVVHCTIQSCPEPQA 1738

RESULT 4  
ID ABB84490 standard; Protein: 1770 AA.  
AC ABB84490;  
DT 24-DEC-2002 (first entry)  
XX Human pregnancy associated protein PAP-Eb SEQ ID 10.  
DE  
XX PAP-E: human; pregnancy associated plasma protein E; abortive;  
KM contraceptive; gene therapy; vaccine; pregnancy; antenatal; diagnosis;  
KM dysgenetic pregnancy.  
XX Homo sapiens.  
OS  
PN US2002102252-A1.  
XX  
XX 01-AUG-2002.  
XX  
XX 06-APR-2001; 2001US-0827998.  
XX  
XX 26-MAY-2000; 2000US-207456P.  
XX  
XX (GUTY/) GU Y.  
XX (SHAN/) SHANNON M E.  
XX  
XX Gu Y, Shannon ME;  
XX  
XX WPI: 2002-697817/75.  
XX N-PSDB; ABS74488; ABS74489.  
XX  
XX New isolated nucleic acid encoding an isoform of human pregnancy  
PT associated plasma protein E, for preventing or aborting pregnancy -  
XX  
XX Claim 8; Figure 4; 353pp; English.  
XX  
XX This invention describes a novel isolated nucleic acid that encodes  
CC one of three new isoforms of human pregnancy associated plasma protein E,  
CC hPAP-E. The products of the invention have abortive and contraceptive  
CC activity and can be used for gene therapy or in a vaccine. The nucleic  
CC acid, polypeptide encoded by it, or antibody to the polypeptide can be  
CC used in pharmaceutical compositions or vaccines for preventing or  
CC aborting pregnancy. PAP-E is used in the antenatal diagnosis of  
CC dysgenetic pregnancies. The nucleic acids are used as probes to assess  
CC the level of PAP-E isoform mRNA in chorionic villus samples, and the  
CC antibodies can be used to assess the expression levels of PAP-E isoform  
CC proteins in chorionic villus samples, to diagnose dysgenetic pregnancies  
CC antenatally. This sequence represents a human PAP-E protein described in  
CC the disclosure of the invention.  
XX  
XX  
SQ Sequence 1770 AA;  
Query Match 85.5%; Score 1532; DB 23; Length 1770;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1732; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 241 QNGEGSYREAEATFNSQVGLPILYFSGRRELLRLREVEIAEIPREAFTEAWMKPEGQON 300  
DB 241 QNGEGSYREAEATFNSQVGLPILYFSGRRELLRLREVEIAEIPREAFTEAWMKPEGQON 300  
QY 301 NPAILIAGVFPNCGSHYSDKMGALGIRSGDKGRDARFPFSLCTDVKATILISHSRQ 360  
DB 301 NPAILIAGVFPNCGSHYSDKMGALGIRSGDKGRDARFPFSLCTDVKATILISHSRQ 360  
QY 361 PGITWHAATYDGRHMAIYVDGTQVASSLDQSPGLSPFMACSRSLGSDSSEDDHYPR 420  
DB 361 PGITWHAATYDGRHMAIYVDGTQVASSLDQSPGLSPFMACSRSLGSDSSEDDHYPR 420  
QY 421 GHLGTLVFMSTALPQSHFOHSSQHSSEEBEATDVLVTAFEPVNTWVPFDEKYPRLV 480  
DB 421 GHLGTLVFMSTALPQSHFOHSSQHSSEEBEATDVLVTAFEPVNTWVPFDEKYPRLV 480  
QY 481 LQGFEPPEPILSLPQPLCGQTCQVCDNVELISQNGVWPLRGENVITYQVNTICDDGLNP 540  
DB 481 LQGFEPPEPILSLPQPLCGQTCQVCDNVELISQNGVWPLRGENVITYQVNTICDDGLNP 540  
QY 541 IYSEBOIRLOHEALNEAFSRXNISWQLSYHQVNSTLRHRVVLVNGEPEKIGNDHCDPEC 600  
DB 541 IYSEBOIRLOHEALNEAFSRXNISWQLSYHQVNSTLRHRVVLVNGEPEKIGNDHCDPEC 600  
QY 601 EHPITGYDGDCHLQGRCYSMNRDGLCHVECNMMLNDFDGDCCDPQVADVAKTCFDDP 660  
DB 601 EHPITGYDGDCHLQGRCYSMNRDGLCHVECNMMLNDFDGDCCDPQVADVAKTCFDDP 660  
QY 661 SPKRAYSVKELKEALQUNSTHFLNITYFASVVEDLAGAATWMDADATHLGCTILSPA 720  
DB 661 SPKRAYSVKELKEALQUNSTHFLNITYFASVVEDLAGAATWMDADATHLGCTILSPA 720  
QY 721 YYGMPGHTDMIEHVGVLGLYVFKGVSEBSCNDPCKETVSMETGDLCACTAATPKS 780  
DB 721 YYGMPGHTDMIEHVGVLGLYVFKGVSEBSCNDPCKETVSMETGDLCACTAATPKS 780  
QY 781 ELCREPEPISDTGCFTRPGAPFTNMYSTDNCTNFPNOVARMHCYLDLVYQWTES 840  
DB 781 ELCREPEPISDTGCFTRPGAPFTNMYSTDNCTNFPNOVARMHCYLDLVYQWTES 840  
QY 841 RKETPIPIPMVIGQNKSLITIMWLPISGVYDRAAGSLCAGCTEDGTRQYVHTASSR 900  
DB 841 RKETPIPIPMVIGQNKSLITIMWLPISGVYDRAAGSLCAGCTEDGTRQYVHTASSR 900  
QY 901 RVCDSSGYMPEEAVGPVDQPCPSLOMSPEVHLVHNMVTPCPTBSCSIELFQHP 960  
DB 901 RVCDSSGYMPEEAVGPVDQPCPSLOMSPEVHLVHNMVTPCPTBSCSIELFQHP 960  
QY 961 VQADTLTLWTSFPMSSQVLPDTEILLENKESVHLGPDLTFCDIPLTIKLHVDKVSQV 1020  
DB 961 VQADTLTLWTSFPMSSQVLPDTEILLENKESVHLGPDLTFCDIPLTIKLHVDKVSQV 1020  
QY 1021 KYTTPDERIRIDALLTSOPHSPLCSGCRPVRYQVLRDPPASGLPVYVTHSRKTTDVE 1080  
DB 1021 KYTTPDERIRIDALLTSOPHSPLCSGCRPVRYQVLRDPPASGLPVYVTHSRKTTDVE 1080  
QY 1081 VTTGQMYQYOVULBAGEGELGEASPLNHIIGAYCGGKXSESLGECDDGDVLVSGDGS 1140  
DB 1081 VTTGQMYQYOVULBAGEGELGEASPLNHIIGAYCGGKXSESLGECDDGDVLVSGDGS 1140  
QY 1141 KVCELEBGFNCVGEPSLCYMYBEDGICEPERKTSIVDCGIYTPKGYLDQMATRAYSHE 1200  
DB 1141 KVCELEBGFNCVGEPSLCYMYBEDGICEPERKTSIVDCGIYTPKGYLDQMATRAYSHE 1200  
QY 1201 DKKKCPVSLVTGPHSLICTSYHPDLPNHRPLTGWFPVCASENETDDREDEQEGSLKE 1260  
DB 1201 DKKKCPVSLVTGPHSLICTSYHPDLPNHRPLTGWFPVCASENETDDREDEQEGSLKE 1260  
QY 1261 DEYWLKVCNRPGEARAFIFLTDTGLVGEHOOPVTVLTVVRGSNHSIGYVJSCQH 1320  
DB 1261 DEYWLKVCNRPGEARAFIFLTDTGLVGEHOOPVTVLTVVRGSNHSIGYVJSCQH 1320  
QY 1321 NPLIINVTTHQNVLFHHTTSVLNLFSSPRVGISAVALTSSRIGLAPSNCISEDEGQNH 1380

Db	Query	Sequence	Score	DB 23	Length	1624
Db	1321	NPILINVTHTQNVLFHHHTTSVLPWFSSPRVIGISAVALTSTRIGLSAPNSCISREDEQNH	1380			
Qy	1381	QGQSCIRHPCGKODSCPSLLIDHADVYNTCTSIGFLMKCAITCGRGFALQASSGQYIRPM	1440			
Db	1381	QGQSCIRHPCGKODSCPSLLIDHADVYNTCTSIGFLMKCAITCGRGFALQASSGQYIRPM	1440			
Qy	1441	QKEILLTCSSGHMWNQNSCLPVDGCVDPSPSLVYNTVANSCSGSGTFGLRCSISCVPAKIQ	1500			
Db	1441	QKEILLTCSSGHMWNQNSCLPVDGCVDPSPSLVYNTVANSCSGSGTFGLRCSISCVPAKIQ	1500			
Qy	1501	GLSPMLTCLIEDGLMSLPBYVCKTECDAPRIILNANLLPHCLQDNHVGITCKYECKRGY	1560			
Db	1501	GLSPMLTCLIEDGLMSLPBYVCKTECDAPRIILNANLLPHCLQDNHVGITCKYECKRGY	1560			
Qy	1561	YVAESAEGKVRNKLKIQCLEGGIMEQSGCIPVCEPPPVFEQMYECTNGFSLDSQCVL	1620			
Db	1561	YVAESAEGKVRNKLKIQCLEGGIMEQSGCIPVCEPPPVFEQMYECTNGFSLDSQCVL	1620			
Qy	1621	NCNOBERKLPILCTKEGLMTQEFLLCENLQCECPPPSELSVYKCEQYIGIGAVCSPL	1680			
Db	1621	NCNOBERKLPILCTKEGLMTQEFLLCENLQCECPPPSELSVYKCEQYIGIGAVCSPL	1680			
Qy	1681	CVIPSPSPVMLPENITADTLEHMMPEPKVOSIVCTGRRQHPDPVLVHCIOQSC	1734			
Db	1681	CVIPSPSPVMLPENITADTLEHMMPEPKVOSIVCTGRRQHPDPVLVHCIOQSC	1734			
RESULT 5						
ABB04485						
ID	ABB04485	standard; Protein; 1624 AA.				
AC	ABB04485;					
DT	14-MAY-2002	(first entry)				
DE	Human pregnancy-associated plasma protein E.					
XX	Human; pregnancy-associated plasma protein E; PAP-E; trisomy 13;					
KM	trisomy 18; trisomy 21; trisomy 22; genetic diagnosis.					
XX	Homo sapiens.					
OS	Homo sapiens.					
XX	WO200195855-A2.					
PN	20-DEC-2001.					
PD	15-JUN-2001; 2001WO-EP06831.					
XX	16-JUN-2000; 2000DE-1028971.					
XX	(TSCH/) TSCHESCHE H.					
XX	(FARR/) FARR M.					
XX	TSchesche H, Farr M;					
XX	WPI; 2002-098007/13.					
XX	N-PSDB; ABA05845.					
XX	New DNA encoding pregnancy-associated plasma protein-E, useful for					
PT	diagnosis of trisomy, also related proteins and antibodies -					
XX	Claim 6; Page 22; 31pp; German.					
XX	The present invention provides the protein and coding sequences of human					
CC	pregnancy-associated plasma protein-E (PAP-E). The sequences can be used					
CC	in human genetic diagnosis, especially for detecting trisomy 13, 18, 21					
CC	or 22. The present sequence is the protein of the invention.					
XX	Sequence 1624 AA;					
XX	Query Match 70.8%; Score 1268; DB 23; Length 1624;					
XX	Best Local Similarity 99.8%; Pred. No. 0;					

Matches	1568:	Conservative	0:	Mismatches	3:	Indels	0:	Gaps	0:
QY	168	MAATTAAITTTLNEKPEPTQRRGMASRRRRQYWKRRADGGQDGSISHHFQWPMHSL	227						
Db	1	MAATTAAITTTLNEKPEPTQRRGMASRRRRQYWKRRADGGQDGSISHHFQWPMHSL	60						
QY	228	KHRVKSPPRESNONGEGSYREAFETNSQVPLVYFSGRRRRLRLRPVLAEIPREAF	287						
Db	61	KHRVKSPPRESNONGEGSYREAFETNSQVPLVYFSGRRRRLRLRPVLAEIPREAF	120						
QY	288	TVEAMVPEEGONNPALIAGVFNDCSHTYVDKGMALGISKGDJKGKDAEFPSLICDRV	347						
Db	121	TVEAMVPEEGONNPALIAGVFNDCSHTYVDKGMALGISKGDJKGKDAEFPSLICDRV	180						
QY	348	KKATILISHSRYPGTWTHVAATYDGHMALVYDGTQVASSLDQSGELNPPMAKCSLL	407						
Db	181	KKATILISHSRYPGTWTHVAATYDGHMALVYDGTQVASSLDQSGELNPPMAKCSLL	240						
QY	408	LGQSSSDGHYFRGHGLTVFWSIALPQSHFQNSQSSSGEEXTDLVLTASPEPVNTEW	467						
Db	241	LGQSSSDGHYFRGHGLTVFWSIALPQSHFQNSQSSSGEEXTDLVLTASPEPVNTEW	300						
QY	468	VPPFDEKYPLEVLVQGEPEPEILSPLOPLCGQTVCDNVELISQYNGVMPRLGEKIRY	527						
Db	301	VPPFDEKYPLEVLVQGEPEPEILSPLOPLCGQTVCDNVELISQYNGVMPRLGEKIRY	360						
QY	528	QVNNICDDEGLNPVSEBOIRLOHEALNEAFSRYNISWQSLVHQVHNSTRHRVLVNC	587						
Db	361	QVNNICDDEGLNPVSEBOIRLOHEALNEAFSRYNISWQSLVHQVHNSTRHRVLVNC	420						
QY	588	PSKIGNDHCPECEHPLTGYDGDCLQGRCYSNRRDGLCHVECNMLNDFDGDCCDP	647						
Db	421	PSKIGNDHCPECEHPLTGYDGDCLQGRCYSNRRDGLCHVECNMLNDFDGDCCDP	480						
QY	648	QVADVTRCTCDPQSPKRAYMSVKEKCALDLSNTHPLNTIYASSVREDLGAATWPMKD	707						
Db	481	QVADVTRCTCDPQSPKRAYMSVKEKCALDLSNTHPLNTIYASSVREDLGAATWPMKD	540						
QY	708	AVTLAGSIVLSPAYYGMGHDTMIHEGVYGLYHYFKVSESESCNDCKETVPMET	767						
Db	541	AVTLAGSIVLSPAYYGMGHDTMIHEGVYGLYHYFKVSESESCNDCKETVPMET	600						
QY	768	GDLCAOTAPTPKSELCREPEPTSDTCGTRFGAPFTNYSYTDNCTDNFTPNQVARM	827						
Db	601	GDLCAOTAPTPKSELCREPEPTSDTCGTRFGAPFTNYSYTDNCTDNFTPNQVARM	660						
QY	828	CYLDLVYQWMTESKRPPIPIPMVYICQTKSLTIHMLPRI SGVYVDRASGSLGACTED	887						
Db	661	CYLDLVYQWMTESKRPPIPIPMVYICQTKSLTIHMLPRI SGVYVDRASGSLGACTED	720						
QY	888	GTFPQVYHTASSRRVCSGTYMPEBAVGPDDVQCEPSLOAMSPEVHL YHNMVTPCP	947						
Db	721	GTFPQVYHTASSRRVCSGTYMPEBAVGPDDVQCEPSLOAMSPEVHL YHNMVTPCP	780						
QY	948	TEGCSLELFOHPVQADTLTLMWTSFPMESSQVLFDEILLENKESVHLGRLDTFCDPI	1007						
Db	781	TEGCSLELFOHPVQADTLTLMWTSFPMESSQVLFDEILLENKESVHLGRLDTFCDPI	840						
QY	1008	TIKLVHDGKYSGVKAVYTFDERIEIDALLTSQHPSLCSCGRPYRYOVLDPAPASGLPV	1067						
Db	841	TIKLVHDGKYSGVKAVYTFDERIEIDALLTSQHPSLCSCGRPYRYOVLDPAPASGLPV	900						
QY	1068	VYTHSHHKFPDVEKTPGOMYOYVLAAGELGASPLNHHNGAPVCSGKXSERIGE	1127						
Db	901	VYTHSHHKFPDVEKTPGOMYOYVLAAGELGASPLNHHNGAPVCSGKXSERIGE	960						
QY	1128	CDDEDLVSGDGSKVCLEBGFNCVSGPSLCYMWEGDGI CEPEFKTYSIVDCGIYTPKGY	1187						
Db	961	CDDEDLVSGDGSKVCLEBGFNCVSGPSLCYMWEGDGI CEPEFKTYSIVDCGIYTPKGY	1020						
QY	1188	LDQWATAYSSHEDKKKCVSLVTGEBHSLICTSYHBDLPNNRPLTGMPCVASENETQD	1247						
Db	1021	LDQWATAYSSHEDKKKCVSLVTGEBHSLICTSYHBDLPNNRPLTGMPCVASENETQD	1080						

QY 1248 DRSEOPGSLKKEDEVMLKVCNRPGEARAFIFLITTDGLVGEHQPTVTLVITDVRGS 1307  
DB 1081 DRSEOPGSLKKEDEVMLKVCNRPGEARAFIFLITTDGLVGEHQPTVTLVITDVRGS 1140  
QY 1308 NNSLGTGSLSCQHNPLIINTVTHQNVLPFHHTTSVLNLFSSPRVGISAVALTSSRIGLSA 1367  
DB 1141 NNSLGTGSLSCQHNPLIINTVTHQNVLPFHHTTSVLNLFSSPRVGISAVALTSSRIGLSA 1200  
QY 1368 PNSCISDEGQNHQSCIRPCGKQSCPSLILDHADVNCISIGGLMKCAITCQGRF 1427  
DB 1201 PNSCISDEGQNHQSCIRPCGKQSCPSLILDHADVNCISIGGLMKCAITCQGRF 1260  
QY 1428 ALQASSGOYIRPMQKEILLTCSSGSHMDQNVSCLPVDCGVPDPSLVNANFSCSEGTFLK 1487  
DB 1261 ALQASSGOYIRPMQKEILLTCSSGSHMDQNVSCLPVDCGVPDPSLVNANFSCSEGTFLK 1320  
QY 1488 RCSIICVPKAKLQGLSFWLTCLDEGLMSLPBVTCKLECDAPITLNNLLPHCLQDND 1547  
DB 1321 RCSIICVPKAKLQGLSFWLTCLDEGLMSLPBVTCKLECDAPITLNNLLPHCLQDND 1380  
QY 1548 VGTICKCKECPGVYVAESAGKVNKLKIQCLEGGIWEQSCIPVCEBPPEVFEQMYE 1607  
DB 1381 VGTICKCKECPGVYVAESAGKVNKLKIQCLEGGIWEQSCIPVCEBPPEVFEQMYE 1440  
QY 1608 CTNGFSLDSQCVLNCNOREKLPILCTKEGLMTOEPKLCENTLOGECPPEPSELNSVEYKC 1667  
DB 1441 CTNGFSLDSQCVLNCNOREKLPILCTKEGLMTOEPKLCENTLOGECPPEPSELNSVEYKC 1500  
QY 1668 EGGYGIGAVCSPLCIVIPSPDPMLENITADTLEHMMBPVKQSVICTGRQWHPDVLV 1727  
DB 1501 EGGYGIGAVCSPLCIVIPSPDPMLENITADTLEHMMBPVKQSVICTGRQWHPDVLV 1560  
QY 1728 HCTOSCEPFOA 1738  
DB 1561 HCTOSCEPFOA 1571

RESULT 6  
ABB84494  
ID ABB84494 standard; Protein: 1385 AA.  
XX ABB84494;  
AC ABB84494;  
DT 24-DEC-2002 (first entry)  
XX  
DE Human pregnancy associated protein PAPP-Ec SEQ ID 16.  
XX  
XX PAPP-E; human; pregnancy associated plasma protein E; abortive;  
KM contraceptive; gene therapy; vaccine; pregnancy; antenatal; diagnosis;  
XX dysgenetic pregnancy.  
XX  
OS Homo sapiens.  
XX  
PN US2002102252-A1.  
XX  
PD 01-AUG-2002.  
XX  
PF 06-APR-2001; 2001US-0827998.  
XX  
PR 26-MAY-2000; 2000US-207456P.  
XX  
PA (GUY/) GU Y.  
XX (SHAN/) SHANNON M E.  
XX  
PI Gu Y, Shannon ME;  
XX  
XX WPI; 2002-697817/75.  
DR N-PSDB; ABS74491.  
XX  
FT New isolated nucleic acid encoding an isoform of human pregnancy  
PT associated plasma protein E, for preventing or aborting pregnancy  
XX

PS Claim 17, Figure 5; 353pp; English.  
XX  
CC This invention describes a novel isolated nucleic acid that encodes  
CC one of three new isoforms of human pregnancy associated plasma protein E,  
CC hPAP-E. The products of the invention have abortive and contraceptive  
CC activity and can be used for gene therapy or in a vaccine. The nucleic  
CC acid, polypeptide encoded by it, or antibody to the polypeptide can be  
CC used in pharmaceutical compositions or vaccines for preventing or  
CC aborting pregnancy. PAPP-E is used in the antenatal diagnosis of  
CC dysgenetic pregnancies. The nucleic acids are used as probes to assess  
CC the level of PAPP-E isoform mRNA in chorionic villus samples, and the  
CC antibodies can be used to assess the expression levels of PAPP-E isoform  
CC proteins in chorionic villus samples, to diagnose dysgenetic pregnancies  
CC antenatally. This sequence represents a human PAPP-E protein described in  
CC the disclosure of the invention.  
XX  
SQ Sequence 1385 AA.  
Query Match 32.0%; Score 574; DB 23; Length 1385;  
Best Local Similarity 99.5%; Pred. No. 0;  
Matches 1074; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 713 GGIVLSPAYYGMGHTDTHIEVGHVGLGYHVKYSERESCHNDPCKETVPSMETDLC 772  
DB 307 GGIVLSPAYYGMGHTDTHIEVGHVGLGYHVKYSERESCHNDPCKETVPSMETDLC 366  
QY 773 DTAPTKSELCREPPTSDPTCGTRPPGAPFTYVMSYTDNCTDNTFPNOVAMHCYLDL 832  
DB 367 DTAPTKSELCREPPTSDPTCGTRPPGAPFTYVMSYTDNCTDNTFPNOVAMHCYLDL 426  
QY 833 VYQWTESRKPTPIPIPMVIGQTNKSLTIHMLPISGVYDRASSLCAGCTEDGTFRQ 892  
DB 427 VYQWTESRKPTPIPIPMVIGQTNKSLTIHMLPISGVYDRASSLCAGCTEDGTFRQ 486  
QY 893 YVHTASSRRVCDSSGWTPEBANGPPDVDPCEPSSIQANSPEVLIYHMMTVCPTEGCS 952  
DB 487 YVHTASSRRVCDSSGWTPEBANGPPDVDPCEPSSIQANSPEVLIYHMMTVCPTEGCS 546  
QY 953 LELLFQHPVADLTLLMTWSFFMESSOVLPTDEILLENKESVHLGPDTRCDPLITIKLH 1012  
DB 547 LELLFQHPVADLTLLMTWSFFMESSOVLPTDEILLENKESVHLGPDTRCDPLITIKLH 606  
QY 1013 VDGKVGSKVYTFDERIEIDALLTSQPHSLCSGCRPVRYQVLRDPPFASGLPVVYTHS 1072  
DB 607 VDGKVGSKVYTFDERIEIDALLTSQPHSLCSGCRPVRYQVLRDPPFASGLPVVYTHS 666  
QY 1073 HRKFTDVEVTPGQMYQYVLAENGSELGASPLNHIHGAIPYCGDGKVSRLGEECDG 1132  
DB 667 HRKFTDVEVTPGQMYQYVLAENGSELGASPLNHIHGAIPYCGDGKVSRLGEECDG 726  
QY 1133 LVSGDGSKVCLEBEGNCVGBSCLCYMBGDCICBPFERKTSIVDCGITYPGYLDQMA 1192  
DB 727 LVSGDGSKVCLEBEGNCVGBSCLCYMBGDCICBPFERKTSIVDCGITYPGYLDQMA 786  
QY 1193 TRAYSSHEDKKKCPVSLVTGEPSHLICTSYHPLPNHRPLTGMFPCVASENETODRSEQ 1252  
DB 787 TRAYSSHEDKKKCPVSLVTGEPSHLICTSYHPLPNHRPLTGMFPCVASENETODRSEQ 846  
QY 1253 PEGSLKKEDEVMLKVCNRPGEARAFIFLITTDGLVGEHQPTVTLVITDVRGSNLSG 1312  
DB 847 PEGSLKKEDEVMLKVCNRPGEARAFIFLITTDGLVGEHQPTVTLVITDVRGSNLSG 906  
QY 1313 TYGLSCQHNPLIINTVTHQNVLPFHHTTSVLNLFSSPRVGISAVALTSSRIGLSABSNCL 1372  
DB 907 TYGLSCQHNPLIINTVTHQNVLPFHHTTSVLNLFSSPRVGISAVALTSSRIGLSABSNCL 966  
QY 1373 SEDEGQNHQSCIRPCGKQSCPSLILDHADVNCISIGGLMKCAITCQGRFALQAS 1432  
DB 967 SEDEGQNHQSCIRPCGKQSCPSLILDHADVNCISIGGLMKCAITCQGRFALQAS 1026  
QY 1433 SGOYIRPMQKEILLTCSSGSHMDQNVSCLPVDCGVPDPSLVNANFSCSEGTFLKRCISIS 1492  
DB 1027 SGOYIRPMQKEILLTCSSGSHMDQNVSCLPVDCGVPDPSLVNANFSCSEGTFLKRCISIS 1086

QY 1493 CVPAPKLGSLPMTLCLEDEGLMSLEPVYCKLECDAPPIILNANLLPHCLQDNHVGFTIC 1552  
DB 1087 CVPAPKLGSLPMTLCLEDEGLMSLEPVYCKLECDAPPIILNANLLPHCLQDNHVGFTIC 1146  
QY 1553 KYECPGYVAESABGKRNKLLKIQCLEGGIWEQSGCIPVCEPPPPVFEQMTCTNGF 1612  
DB 1147 KYECPGYVAESABGKRNKLLKIQCLEGGIWEQSGCIPVCEPPPPVFEQMTCTNGF 1206  
QY 1613 SLDSQCVLNCNQEERKLPILCTKSGLMTOEERLCENLQGEPPPSSEINSEVYCEQGYG 1672  
DB 1207 SLDSQCVLNCNQEERKLPILCTKSGLMTOEERLCENLQGEPPPSSEINSEVYCEQGYG 1266  
QY 1673 IGAVCSPLCVIPSPDPVMLPENITADTLEHMEPEVKQSIYCTGRGRQHPDPVLVHCIOG 1732  
DB 1267 IGAVCSPLCVIPSPDPVMLPENITADTLEHMEPEVKQSIYCTGRGRQHPDPVLVHCIOG 1326  
QY 1733 CEPFOADGWCCTINNRAYCHYDGGDCSSSTLSSKKVIPPADCDLDECTCRDPKAEENQ 1791  
DB 1327 CEPFOADGWCCTINNRAYCHYDGGDCSSSTLSSKKVIPPADCDLDECTCRDPKAEENQ 1385

RESULT 7  
AAM24060  
ID AAM24060 standard; Protein; 317 AA.  
XX AAM24060;  
AC  
XX  
DT 12-OCT-2001 (first entry)  
XX  
DE Human EST encoded protein SEQ ID NO: 1585.  
XX  
KM Human; sheep; pig; cow; fruit fly; Yeast; hamster; macaque; horse;  
KM tomato; monkey; dog; sea urchin; expressed sequence tag; EST;  
KM diagnostics; forensic test; gene mapping; genetic disorder;  
KM biodiversity; gene therapy; nutrition.  
XX  
OS Homo sapiens.  
XX  
PN WO200154477-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 25-JAN-2001; 2001WO-US02687.  
XX  
PR 25-JAN-2000; 2000US-0491404.  
PR 17-JUL-2000; 2000US-0617746.  
PR 03-AUG-2000; 2000US-0631451.  
PR 15-SEP-2000; 2000US-0663870.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V,  
PI Cao Y, Dmanac RA, Zhang J, Wehman T;  
XX  
XX WPI; 2001-476164/51.  
DR N-PSDB; AAH98719.  
XX  
XX Isolated polypeptide for treatment of diseases, diagnostics, raising  
PT antibodies and research use -  
XX  
XX Claim 20; Page 1084; 1275pp; English.  
XX  
XX The present invention provides the protein and coding sequences of novel  
CC proteins from a variety of organisms, including human, dog, cat, horse,  
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea  
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)  
CC from the organism of interest. They can be used in diagnostics,  
CC forensics, gene mapping, identification of mutations, to assess  
CC biodiversity and for nutritional purposes. The present sequence is a  
CC protein of the invention.  
XX  
SQ Sequence 317 AA;

Query Match 17.1%; Score 307; DB 22; Length 317;  
Best Local Similarity 100.0%; Pred. No. 1.9e-302;  
Matches 307; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MCKILIRISLAILAGALCSANSELGWTNRKSLVEREHLNOVLLBGERCWLGAQRPR 60  
DB 1 MCKILIRISLAILAGALCSANSELGWTNRKSLVEREHLNOVLLBGERCWLGAQRPR 60  
QY 61 ASPQHLLFGYVPSAAGVYLRPYRGEQIHTHTGSKSDPTDEGNVSLVPPDLTENPAGIRG 120  
DB 61 ASPQHLLFGYVPSAAGVYLRPYRGEQIHTHTGSKSDPTDEGNVSLVPPDLTENPAGIRG 120  
QY 121 AVEEPAPWQDSPIQSGSELLGDDAYLGNORSKESLGEAGIQGSMMAATTTAIFFTL 180  
DB 121 AVEEPAPWQDSPIQSGSELLGDDAYLGNORSKESLGEAGIQGSMMAATTTAIFFTL 180  
QY 181 NEPRPETQRGMWAKSRQRQVWKRRAEDGQDSGISHPQWPRHSLKRVKSPPEESN 240  
DB 181 NEPRPETQRGMWAKSRQRQVWKRRAEDGQDSGISHPQWPRHSLKRVKSPPEESN 240  
QY 241 QNGEGSYREAFETNSQVGLPIYFSGRERLLRPEVLAIPREAFVEMVYKPEEGQN 300  
DB 241 QNGEGSYREAFETNSQVGLPIYFSGRERLLRPEVLAIPREAFVEMVYKPEEGQN 300  
QY 301 NPATIAAG 307  
DB 301 NPATIAAG 307

RESULT 8  
ABG48354  
ID ABG48354 standard; Peptide; 192 AA.  
XX  
XX ABG48354;  
AC  
XX  
DT 25-FEB-2003 (first entry)  
XX  
DE Human liver peptide, SEQ ID No 27002.  
XX  
KM Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;  
KM hypercholesterolaemia; coronary heart disease.  
XX  
OS Homo sapiens.  
XX  
PN WO200157273-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US00664.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2001-488898/53.  
DR  
XX Human genome-derived single exon nucleic acid probes useful for  
PT analysing gene expression in human adult liver -  
XX  
XX Claim 27; SEQ ID No 27002; 658pp; English.  
XX  
XX The invention relates to a single exon nucleic acid probe (SENP) (I) for  
CC measuring human gene expression in a sample derived from human adult  
CC liver, comprising one of 13109 defined nucleotide sequences given in the

CC specification (or complements/ fragments). The probe hybridises at high  
CC stringency to a nucleic acid molecule expressed in the human adult  
CC liver. (1) may be used for predicting, measuring and displaying gene  
CC expression in samples derived from human adult liver. The genes  
CC identified may be involved in genetic liver diseases such as cirrhosis,  
CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which  
CC is associated with coronary heart disease. ABG47348-ABG59930 represent  
CC human liver single exon encoded peptides of the invention.  
CC Note: The sequence information for this patent does not appear in the  
CC printed specification but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

CC Sequence 192 AA;

Query Match 10.7%; Score 192; DB 22; Length 192;

Best Local Similarity 100.0%; Pred. No. 7.9e-186;

Matches 192; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 472 DEKYRLVLOGFEEPEPILSPLOPPLCGQTVCDNVELISQYNGWPLRGEKVIRYQVN 531  
DB 1 DEKYRLVLOGFEEPEPILSPLOPPLCGQTVCDNVELISQYNGWPLRGEKVIRYQVN 60  
QY 532 ICDDGGLNPVISEEQIRLOHEALNEAFSRVNI SWQSVHQVNSTLRHRVVLVNCPEPSKI 591  
DB 61 ICDDGGLNPVISEEQIRLOHEALNEAFSRVNI SWQSVHQVNSTLRHRVVLVNCPEPSKI 120  
QY 532 GNDHCDPECEHPLTGYDGGDRLQGRCYSMNRDGLCHECNMMLNDPDDGCCDPOVAD 651  
DB 121 GNDHCDPECEHPLTGYDGGDRLQGRCYSMNRDGLCHECNMMLNDPDDGCCDPOVAD 180  
QY 652 VRKTCFDPDSPK 663  
DB 181 VRKTCFDPDSPK 192

RESULT 9

ABB28331 standard; Peptide; 192 AA.

ID ABB28331;

DT 01-FEB-2002 (first entry)

DE Human peptide #982 encoded by breast cell single exon nucleic acid probe.

XX Human; microarray; single exon probe; gene expression; breast;

KM disease; cancer.

XX Homo sapiens.

OS WO200157271-A2.

PN 09-AUG-2001.

PD 30-JAN-2001; 2001WO-US00662.

PF 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

PR 04-OCT-2000; 2000GB-0024263.

PR 04-OCT-2000; 2000GB-0024263.

PR 04-OCT-2000; 2000GB-0024263.

PR 04-OCT-2000; 2000GB-0024263.

PR 04-OCT-2000; 2000GB-0024263.

PR 04-OCT-2000; 2000GB-0024263.

PR 04-OCT-2000; 2000GB-0024263.

PR 04-OCT-2000; 2000GB-0024263.

PR 04-OCT-2000; 2000GB-0024263.

PR 04-OCT-2000; 2000GB-0024263.

PR 04-OCT-2000; 2000GB-0024263.

PR 04-OCT-2000; 2000GB-0024263.

PR 04-OCT-2000; 2000GB-0024263.

PR 04-OCT-2000; 2000GB-0024263.

PR 04-OCT-2000; 2000GB-0024263.

XX Claim 27; SEQ ID NO 11299; 327bp + sequence listing; English.

PS The invention relates to a spatially-addressable set of single exon  
XX nucleic acid probes for measuring gene expression in a sample derived  
XX from human breast and BT 474 cells. The method involves contacting  
XX the probes with a collection of detectably labelled nucleic acids  
XX derived from mRNA of human breast, and then measuring the label  
XX bound to each probe of the microarray. The probes are useful for  
XX verifying the expression of regions of genomic DNA predicted to  
XX encode proteins. They are useful for gene discovery, and for  
XX determining predisposition and/or prognosing breast disease. Gene  
XX expression analysis is useful for assessing the toxicity of chemical  
XX agents on cells. The microarray of this invention presents a far greater  
XX diversity of probes for measuring gene expression, with far less bias  
XX than expressed sequence tag microarrays. The method is suitable for  
XX rapid production of functional information from genomic sequence. The  
XX present sequence is a peptide encoded by a single exon nucleic acid  
XX probe of the invention.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

CC Sequence 192 AA;

Query Match 10.7%; Score 192; DB 22; Length 192;

Best Local Similarity 100.0%; Pred. No. 7.9e-186;

Matches 192; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 472 DEKYRLVLOGFEEPEPILSPLOPPLCGQTVCDNVELISQYNGWPLRGEKVIRYQVN 531  
DB 1 DEKYRLVLOGFEEPEPILSPLOPPLCGQTVCDNVELISQYNGWPLRGEKVIRYQVN 60  
QY 532 ICDDGGLNPVISEEQIRLOHEALNEAFSRVNI SWQSVHQVNSTLRHRVVLVNCPEPSKI 591  
DB 61 ICDDGGLNPVISEEQIRLOHEALNEAFSRVNI SWQSVHQVNSTLRHRVVLVNCPEPSKI 120  
QY 532 GNDHCDPECEHPLTGYDGGDRLQGRCYSMNRDGLCHECNMMLNDPDDGCCDPOVAD 651  
DB 121 GNDHCDPECEHPLTGYDGGDRLQGRCYSMNRDGLCHECNMMLNDPDDGCCDPOVAD 180  
QY 652 VRKTCFDPDSPK 663  
DB 181 VRKTCFDPDSPK 192

RESULT 10

ABB33508 standard; Peptide; 192 AA.

ID ABB33508;

DT 04-FEB-2002 (first entry)

DE Peptide #1014 encoded by human foetal liver single exon probe.

XX Human; foetal liver; gene expression; single exon nucleic acid probe.

KM Homo sapiens.

OS WO200157277-A2.

PN 09-AUG-2001.

PD 30-JAN-2001; 2001WO-US00669.

PF 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

PR 04-OCT-2000; 2000GB-0024263.

PR 04-OCT-2000; 2000GB-0024263.

PR 04-OCT-2000; 2000GB-0024263.

PR 04-OCT-2000; 2000GB-0024263.

PR 04-OCT-2000; 2000GB-0024263.

PR 04-OCT-2000; 2000GB-0024263.

PR 04-OCT-2000; 2000GB-0024263.

PR 04-OCT-2000; 2000GB-0024263.

PR 04-OCT-2000; 2000GB-0024263.

PR 04-OCT-2000; 2000GB-0024263.



XX (MOLE-) MOLECULAR DYNAMICS INC.  
 PA Penn SG, Hanzel DK, Chen W, Rank DR;  
 PI WPI; 2001-483447/52.  
 DR  
 XX Human genome-derived single exon nucleic acid probes useful for  
 PT analyzing gene expression in human fetal liver -  
 XX  
 PS Claim 27; SEQ ID NO 26143; 639bp + sequence listing; English.  
 CC The invention relates to a single exon nucleic acid probe for  
 CC measuring human gene expression in a sample derived from human foetal  
 CC liver. The single exon nucleic acid probes may be used for predicting,  
 CC measuring and displaying gene expression in samples derived from human  
 CC fetal liver. The present sequence is a peptide encoded by a single exon  
 CC nucleic acid probe of the invention.  
 CC Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 XX  
 SQ Sequence 192 AA;  
 Query Match 10.7%; Score 192; DB 22; Length 192;  
 Best Local Similarity 100.0%; Pred. No. 7.9e-186;  
 Matches 192; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 472 DEKPRLEVLQGFPEPEPILSPLOPLCGQIVCNVELISQYNGWPLRGEKVIIRYQVYN 531  
 Db 1 DEKPRLEVLQGFPEPEPILSPLOPLCGQIVCNVELISQYNGWPLRGEKVIIRYQVYN 60  
 QY 532 ICDEGLNPVISEQIRLQHEALNEAFSRVNIISWQLSVHGVHNSLTRRVLVYNCEPSKI 591  
 Db 61 ICDEGLNPVISEQIRLQHEALNEAFSRVNIISWQLSVHGVHNSLTRRVLVYNCEPSKI 120  
 QY 592 GNDHCDPECEHPLTGYDGDGCRLOGRCYSWNRDGLCHEVCNNMLNDFDDGCCDPQVAD 651  
 Db 121 GNDHCDPECEHPLTGYDGDGCRLOGRCYSWNRDGLCHEVCNNMLNDFDDGCCDPQVAD 180  
 QY 652 VRKTCFDPDSPK 663  
 Db 181 VRKTCFDPDSPK 192  
 RESULT 11  
 ABB18967  
 ID ABB18967 standard; Protein; 192 AA.  
 XX  
 AC ABB18967;  
 XX  
 DT 23-JAN-2002 (first entry)  
 XX  
 DE Protein #966 encoded by probe for measuring heart cell gene expression.  
 XX  
 KW Human: gene expression; heart; microarray; vascular system;  
 KW cardiovascular disease; hypertension; cardiac arrhythmia;  
 KW congenital heart disease.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157274-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 30-JAN-2001; 2001WO-US00666.  
 XX  
 PR 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 XX (MOLE-) MOLECULAR DYNAMICS INC.  
 PA Penn SG, Hanzel DK, Chen W, Rank DR;  
 PI WPI; 2001-488899/53.  
 DR  
 XX Single exon nucleic acid probes for analyzing gene expression in human  
 PT hearts -  
 XX  
 PS Claim 15; SEQ ID NO 20737; 530bp; English.  
 CC The present invention relates to single exon nucleic acid probes for  
 CC measuring human gene expression in a sample derived from human heart (see  
 CC ABA21535-ABA41305). The present sequence is a protein encoded by one such  
 CC probe. The probes may be used for predicting, measuring and displaying  
 CC gene expression in samples derived from the human heart via microarrays.  
 CC By measuring gene expression, the probes are useful for predicting,  
 CC diagnosing, grading, staging, monitoring and prognosing diseases of the  
 CC human heart and vascular system e.g. cardiovascular disease.  
 CC hypertension, cardiac arrhythmias and congenital heart disease.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 XX  
 SQ Sequence 192 AA;  
 Query Match 10.7%; Score 192; DB 22; Length 192;  
 Best Local Similarity 100.0%; Pred. No. 7.9e-186;  
 Matches 192; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 472 DEKPRLEVLQGFPEPEPILSPLOPLCGQIVCNVELISQYNGWPLRGEKVIIRYQVYN 531  
 Db 1 DEKPRLEVLQGFPEPEPILSPLOPLCGQIVCNVELISQYNGWPLRGEKVIIRYQVYN 60  
 QY 532 ICDEGLNPVISEQIRLQHEALNEAFSRVNIISWQLSVHGVHNSLTRRVLVYNCEPSKI 591  
 Db 61 ICDEGLNPVISEQIRLQHEALNEAFSRVNIISWQLSVHGVHNSLTRRVLVYNCEPSKI 120  
 QY 592 GNDHCDPECEHPLTGYDGDGCRLOGRCYSWNRDGLCHEVCNNMLNDFDDGCCDPQVAD 651  
 Db 121 GNDHCDPECEHPLTGYDGDGCRLOGRCYSWNRDGLCHEVCNNMLNDFDDGCCDPQVAD 180  
 QY 652 VRKTCFDPDSPK 663  
 Db 181 VRKTCFDPDSPK 192  
 RESULT 12  
 AAM54287  
 ID AAM54287 standard; Protein; 192 AA.  
 XX  
 AC AAM54287;  
 XX  
 DT 05-NOV-2001 (first entry)  
 XX  
 DE Human brain expressed single exon probe encoded protein SEQ ID NO: 26392.  
 XX  
 KW Human; brain expressed exon; gene expression analysis; probe;  
 KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
 KW epilepsy; cancer.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157275-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 30-JAN-2001; 2001WO-US00667.  
 XX  
 PR 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-483446/52.  
XX  
XX Single exon nucleic acid probes for analyzing gene expression in human  
XX brains -  
XX  
XX Example 4; SEQ ID NO: 26392; 650bp + Sequence Listing; English.  
XX  
XX The present invention provides a number of single exon nucleic acid  
XX probes which are derived from genomic sequences expressed in the human  
XX brain. They can be used to measure gene expression in brain cell samples,  
XX which may enable the diagnosis and improved treatment of nervous system  
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
XX epilepsy and cancers. The present sequence is a protein encoded by one of  
XX the probes of the invention.  
XX  
XX Sequence 192 AA;  
SQ

Query Match 10.7%; Score 192; DB 22; Length 192;  
Best Local Similarity 100.0%; Pred. No. 7.9e-186;  
Matches 192; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 472 DEKYRLEVLQGFEPPEILSPLOPLCGQTVCDNVELISQYNGWPLRGEKVIRYQVN 531  
DB 1 DEKYRLEVLQGFEPPEILSPLOPLCGQTVCDNVELISQYNGWPLRGEKVIRYQVN 60  
QY 532 ICDEGLNPVISEQIRLOHEALNEAFSRYNISWQLSVQVNSTLRHVLVNCEPSKI 591  
DB 61 ICDEGLNPVISEQIRLOHEALNEAFSRYNISWQLSVQVNSTLRHVLVNCEPSKI 120  
QY 592 GNDHCDPECEHPPLTGYDGDGCRLOGRCYSMNRDGLCHECNMNLNDFDDGCCDPQVAD 651  
DB 121 GNDHCDPECEHPPLTGYDGDGCRLOGRCYSMNRDGLCHECNMNLNDFDDGCCDPQVAD 180  
QY 652 VRKTCFDPDSPK 663  
DB 181 VRKTCFDPDSPK 192

## RESULT 13

AA66682  
ID AAM66682 standard; Protein; 192 AA.  
XX  
XX AAM66682;  
XX  
XX DT 06-NOV-2001 (first entry)  
XX  
XX Human bone marrow expressed probe encoded protein SEQ ID NO: 26388.  
XX  
XX Human; bone marrow expressed exon; gene expression analysis; probe;  
XX microarray; cancer; leukaemia; lymphoma; myeloma.  
XX  
XX Homo sapiens.  
XX  
XX WO200157276-A2.  
XX  
XX PD 09-AUG-2001.  
XX  
XX PF 30-JAN-2001; 2001WO-US00668.  
XX  
XX PR 04-FEB-2000; 2000US-0180312.  
XX PR 26-MAY-2000; 2000US-0207456.  
XX PR 30-JUN-2000; 2000US-0608408.  
XX PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-488900/53.  
XX  
XX Human genome-derived single exon nucleic acid probes useful for  
XX analyzing gene expression in human bone marrow -  
XX  
XX Example 4; SEQ ID NO: 26988; 658bp + Sequence Listing; English.  
XX  
XX The present invention provides a number of single exon nucleic acid  
XX probes which are derived from genomic sequences expressed in the human  
XX bone marrow. They can be used to measure gene expression in bone marrow  
XX samples, which may enable the improved diagnosis and treatment of cancers  
XX such as lymphoma, leukaemia and myeloma. The present sequence is a  
XX protein encoded by one of the probes of the invention.  
XX  
XX Sequence 192 AA;  
SQ

Query Match 10.7%; Score 192; DB 22; Length 192;  
Best Local Similarity 100.0%; Pred. No. 7.9e-186;  
Matches 192; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 472 DEKYRLEVLQGFEPPEILSPLOPLCGQTVCDNVELISQYNGWPLRGEKVIRYQVN 531  
DB 1 DEKYRLEVLQGFEPPEILSPLOPLCGQTVCDNVELISQYNGWPLRGEKVIRYQVN 60  
QY 532 ICDEGLNPVISEQIRLOHEALNEAFSRYNISWQLSVQVNSTLRHVLVNCEPSKI 591  
DB 61 ICDEGLNPVISEQIRLOHEALNEAFSRYNISWQLSVQVNSTLRHVLVNCEPSKI 120  
QY 592 GNDHCDPECEHPPLTGYDGDGCRLOGRCYSMNRDGLCHECNMNLNDFDDGCCDPQVAD 651  
DB 121 GNDHCDPECEHPPLTGYDGDGCRLOGRCYSMNRDGLCHECNMNLNDFDDGCCDPQVAD 180  
QY 652 VRKTCFDPDSPK 663  
DB 181 VRKTCFDPDSPK 192

## RESULT 14

AAM14550  
ID AAM14550 standard; Protein; 192 AA.  
XX  
XX AAM14550;  
XX  
XX DT 12-OCT-2001 (first entry)  
XX  
XX Peptide #984 encoded by probe for measuring cervical gene expression.  
XX Probe; human; microarray; gene expression; cervical epithelial cell;  
XX cervical cancer.  
XX  
XX Homo sapiens.  
XX  
XX WO200157278-A2.  
XX  
XX PD 09-AUG-2001.  
XX  
XX PF 30-JAN-2001; 2001WO-US00670.  
XX  
XX PR 04-FEB-2000; 2000US-0180312.  
XX PR 26-MAY-2000; 2000US-0207456.  
XX PR 30-JUN-2000; 2000US-0608408.  
XX PR 03-AUG-2000; 2000US-0632366.  
XX PR 21-SEP-2000; 2000US-0234687.  
XX PR 27-SEP-2000; 2000US-0236359.  
XX PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.  
PA Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-488901/53.  
XX Human genome-derived single exon nucleic acid probes useful for  
PT analyzing gene expression in human cervical epithelial cells -  
XX  
XX Claim 27; SEQ ID No 19376; 487bp; English.  
XX  
XX The present invention relates to human single exon nucleic acid probes  
CC (SRNP: see A110068-A128459). The present sequence is a peptide encoded  
CC by one such probe. The SENPs are derived from human HeLa cells. The SENPs  
CC can be used to produce a single exon microarray, which can be used for  
CC measuring human gene expression in a sample derived from human cervical  
CC epithelial cells. By measuring gene expression, the probes are therefore  
CC useful in grading and/or staging of diseases of the cervix, notably  
CC cervical cancer.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pcc\_sequences.  
XX  
SQ Sequence 192 AA;  
Query Match 10.7%; Score 192; DB 22; Length 192;  
Best Local Similarity 100.0%; Pred. No. 7.9e-186;  
Matches 192; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 472 DEKYPRLVLOGFEPPEPILSPLOPLCGQVCDNVELISQYNGWPLRGEKVIIRYQVN 531  
DB 1 DEKYPRLVLOGFEPPEPILSPLOPLCGQVCDNVELISQYNGWPLRGEKVIIRYQVN 60  
QY 532 ICDEGLNPIVSEEQIRLOHEALNEAFSRVNI SMQLSVHGVNSTLRHRVYL VNCPEPSKI 591  
DB 61 ICDEGLNPIVSEEQIRLOHEALNEAFSRVNI SMQLSVHGVNSTLRHRVYL VNCPEPSKI 120  
QY 592 GNDHCDPECEHPLTGYDGGDCRLQGRCYSMNRDGLCHVECNMMLNDFDDGCCDPQYAD 651  
DB 121 GNDHCDPECEHPLTGYDGGDCRLQGRCYSMNRDGLCHVECNMMLNDFDDGCCDPQYAD 180  
QY 652 VRKTCFDPDSPK 663  
DB 181 VRKTCFDPDSPK 192

RESULT 15  
AAM26968  
ID AAM26968 standard; Protein; 192 AA.  
XX  
XX AAM26968;  
XX  
DT 17-OCT-2001 (first entry)  
XX  
XX Peptide #1005 encoded by probe for measuring placental gene expression.  
DE  
XX Probe; microarray; human; placenta; antenatal diagnosis;  
KW genetic disorder.  
XX  
XX Homo sapiens.  
OS  
XX WO200157272-A2.  
PN  
XX 09-AUG-2001.  
PD  
XX 30-JAN-2001; 2001WO-US00663.  
PF  
XX 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
PA Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-48897/53.  
XX Human genome-derived single exon nucleic acid probes useful for  
PT analyzing gene expression in human placenta -  
XX  
XX Claim 27; SEQ ID No 27237; 654bp; English.  
XX  
XX The present invention relates to single exon nucleic acid probes (SENPs;  
CC see A133115-A157546). The present sequence is a peptide encoded by one  
CC such probe. The probes are useful for producing a microarray for  
CC predicting, measuring and displaying gene expression in samples derived  
CC from human placenta. The probes are useful for antenatal diagnosis of  
CC human genetic disorders.  
XX  
SQ Sequence 192 AA;  
Query Match 10.7%; Score 192; DB 22; Length 192;  
Best Local Similarity 100.0%; Pred. No. 7.9e-186;  
Matches 192; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 472 DEKYPRLVLOGFEPPEPILSPLOPLCGQVCDNVELISQYNGWPLRGEKVIIRYQVN 531  
DB 1 DEKYPRLVLOGFEPPEPILSPLOPLCGQVCDNVELISQYNGWPLRGEKVIIRYQVN 60  
QY 532 ICDEGLNPIVSEEQIRLOHEALNEAFSRVNI SMQLSVHGVNSTLRHRVYL VNCPEPSKI 591  
DB 61 ICDEGLNPIVSEEQIRLOHEALNEAFSRVNI SMQLSVHGVNSTLRHRVYL VNCPEPSKI 120  
QY 592 GNDHCDPECEHPLTGYDGGDCRLQGRCYSMNRDGLCHVECNMMLNDFDDGCCDPQYAD 651  
DB 121 GNDHCDPECEHPLTGYDGGDCRLQGRCYSMNRDGLCHVECNMMLNDFDDGCCDPQYAD 180  
QY 652 VRKTCFDPDSPK 663  
DB 181 VRKTCFDPDSPK 192

Search completed: January 2, 2004, 16:13:08  
Job time : 56 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2004, 16:11:13 ; Search time 29 Seconds  
(without alignments)  
5939.245 Million cell updates/sec

Title: US-09-983-025A-2

Perfect score: 1791

Sequence: 1 MMCKIIRISLIALAGWALC.....AADCDDECTCRDPKAEENQ 1791

Scoring table: OLIGO

Searched: 283308 seqs, 96168682 residues

Word size: 0

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

1: PIR\_76:.\*  
2: PIR1:.\*  
3: PIR3:.\*  
4: PIR4:.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18	1.0	1627	2	S65464 pregnancy-associated
2	9	0.5	502	2	D86765 hypothetical prote
3	8	0.4	105	1	S40199 ribosomal protein
4	8	0.4	252	2	AE2440 hypothetical prote
5	8	0.4	349	2	P95887 probable ABC trans
6	8	0.4	377	2	T04089 GTP-binding protei
7	8	0.4	407	2	B71262 probable glutamate
8	8	0.4	422	1	BVBKMS M511 protein - yea
9	8	0.4	457	2	D95194 hypothetical prote
10	8	0.4	457	2	A98061 UDP-N-acetylmuramo
11	8	0.4	467	2	T17174 hypothetical prote
12	8	0.4	538	2	S62566 p1D finger protein
13	8	0.4	580	2	B86177 hypothetical prote
14	8	0.4	592	2	UC4642 purH bifunctional
15	8	0.4	596	2	A28088 oxalacetate decar
16	8	0.4	620	2	AF2174 hypothetical prote
17	8	0.4	712	2	AG0175 probable membrane
18	8	0.4	1638	2	T30313 chemotaxis protein
19	8	0.4	2472	2	B83594 still framehift p
20	8	0.4	4563	1	LPHUB apolipoprotein B-1
21	7	0.4	75	2	H69273 hypothetical prote
22	7	0.4	93	2	E44196 orf5 protein - Mar
23	7	0.4	98	2	D96797 Sm-like protein (i
24	7	0.4	103	1	G64093 ribosomal protein
25	7	0.4	103	1	RSB824 ribosomal protein
26	7	0.4	103	1	RSB828 ribosomal protein
27	7	0.4	103	1	T44394 ribosomal protein
28	7	0.4	104	1	R5EC24 ribosomal protein
29	7	0.4	104	2	JC2277 ribosomal protein

30	7	0.4	104	2	F91150 50S ribosomal subu
31	7	0.4	104	2	B83115 50S ribosomal prot
32	7	0.4	104	2	B85996 50S ribosomal subu
33	7	0.4	104	2	AE0027 50S ribosomal prot
34	7	0.4	104	2	AD1007 50S ribosomal chai
35	7	0.4	107	2	D83918 hypothetical prote
36	7	0.4	113	2	S29883 ribosomal protein
37	7	0.4	115	2	S77491 ribosomal protei
38	7	0.4	120	2	F72450 hypothetical prote
39	7	0.4	122	2	B87622 response regulator
40	7	0.4	125	2	C70637 hypothetical prote
41	7	0.4	136	2	AG2590 conserved hypothe
42	7	0.4	142	2	F86443 unknown protein (i
43	7	0.4	147	1	FESP1 ferridoxin (2Fe-2S
44	7	0.4	154	2	F70312 riboflavin synthas
45	7	0.4	155	2	S27018 bactericin precurs

#### ALIGNMENTS

RESULT 1  
S65464 pregnancy-associated plasma protein A precursor - human  
N:Alternate names: PAP-A  
C:Species: Homo sapiens (man)  
C:Date: 22-Nov-1996 #sequence revision 22-Nov-1996 #text\_change 05-Nov-1999  
C:Accession: S65464; S65463; A54220; I38097  
R:Haaning, J.; Oxvig, C.; Overgaard, M.T.; Ebbesen, P.; Kristensen, T.; Sottrup-Jensen, I.  
submitted to the EMBL Data Library, June 1995  
A:Description: Complete cDNA sequence of the preproform of human pregnancy-associated plasma  
A:Reference number: S65464  
A:Accession: S65464  
A:Molecule type: mRNA  
A:Residues: 1-1627 <HAA>  
A:Cross-references: EMBL:U28727; NID:G1142969; PIDN:AAC50543.1; PID:G1142970  
R:Haaning, J.; Oxvig, C.; Overgaard, M.T.; Ebbesen, P.; Kristensen, T.; Sottrup-Jensen, I.  
Eur. J. Biochem. 237, 159-163, 1996  
A:Title: Complete cDNA sequence of the preproform of human pregnancy-associated plasma  
A:Reference number: S65463; MUID:96203921; PMID:8620868  
A:Accession: S65463  
A:Molecule type: mRNA  
A:Residues: 1-102 <HAM>  
A:Cross-references: EMBL:U28727  
A:Note: the authors translated the codon CGA for residue 101 as Thr  
R:Kristensen, T.; Oxvig, C.; Sand, O.; Moller, N.P.H.; Sottrup-Jensen, L.  
Biochemistry 33, 1592-1598, 1994  
A:Title: Amino acid sequence of human pregnancy-associated plasma protein-A derived from  
A:Reference number: A54220; MUID:94146014; PMID:7508748  
A:Accession: A54220  
A:Molecule type: mRNA  
A:Residues: 77-1627 <KRI>  
A:Cross-references: GB:X68280; NID:G394649; PIDN:CAA6341.1; PID:G394650  
R:Oxvig, C.; Sand, O.; Kristensen, T.; Gleich, G.J.; Sottrup-Jensen, L.  
J. Biol. Chem. 268, 12243-12246, 1993  
A:Title: Circulating human pregnancy-associated plasma protein-A is disulfide-bridged to  
A:Reference number: I38097; MUID:93286045; PMID:7685339  
A:Accession: I38097  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 77-1627 <RES>  
A:Cross-references: EMBL:X68280; NID:G394649; PIDN:CAA6341.1; PID:G394650  
C:Genetic:1  
A:Gene: GDB:PAPPA  
A:Cross-references: GDB:134729; OMIM:176385  
A:Map position: 9q33.1-9q33.1  
F:1-82/Domain: signal sequence #status predicted <SIG>  
F:23-60/Domain: propeptide #status predicted <PRO>  
F:81-1627/Product: pregnancy-associated plasma protein A #status predicted <MAT>

Query Match 1.0%; Score 18; DB 2; Length 1627;  
Best Local Similarity 100.0%; Pred. No. 3.1e-09;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 818 FTNQNARMHCYLDLVYQ 835  
|||  
Db 647 FTNQNARMHCYLDLVYQ 664

## RESULT 2

D86765  
hypothetical protein frdc [imported] - Lactococcus lactis subsp. lactis (strain IL1403)  
C:Species: Lactococcus lactis subsp. lactis  
C:Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 03-Aug-2001  
C:Accession: D86765  
R:Botolin, A.; Wincker, P.; Manger, S.; Jallion, O.; Malarme, K.; Weissenbach, J.; Ehrlich  
Genome Res. 11, 731-753, 2001  
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis s  
A:Reference number: A86625; MUID:21235186; PMID:11337471  
A:Accession: D86765  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-502 <STD>  
A:Cross-references: GB:AE005176; PID:g12724085; PIDN:AAK05222.1; GSPDB:GN00146  
C:Genetics:  
A:Gene: frdc  
C:Superfamily: osmotic growth protein 1; 3-oxosteroid 1-dehydrogenase homology; fumarate

Query Match 0.5%; Score 9; DB 2; Length 502;  
Best Local Similarity 100.0%; Pred. No. 2.8;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1014 DGKVGCVK 1022  
|||  
Db 217 DGKVGCVK 225

## RESULT 3

S40199  
ribosomal protein L24 - Thermotoga maritima (strain MSB8)  
C:Species: Thermotoga maritima  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 21-Jul-2000  
C:Accession: S40199; A72249  
R:Sanangelantoni, A.; Tidoni, O.  
submitted to the EMBL Data Library, February 1993  
A:Reference number: S37489  
A:Accession: S40199  
A:Molecule type: DNA  
A:Residues: 1-105 <SAN>  
A:Cross-references: EMBL:Z21677; NID:g437921; PIDN:CAA79788.1; PID:g437934  
R:Neilson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey  
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;  
C.M.  
Nature 399, 323-329, 1999  
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq  
A:Reference number: A72200; MUID:99287316; PMID:10360571  
A:Accession: A72249  
A:Molecule type: DNA  
A:Residues: 1-105 <RAN>  
A:Cross-references: GB:AE001798; GB:AE000512; NID:g4982033; PIDN:AA036555.1; PID:g498205  
C:Genetics:  
A:Gene: TM1489  
A:Superfamily: Escherichia coli ribosomal protein L24  
C:Keywords: protein biosynthesis; ribosome

Query Match 0.4%; Score 8; DB 1; Length 105;  
Best Local Similarity 100.0%; Pred. No. 8.5;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 327 SGKDKGR 334  
|||  
Db 13 SGKDKGR 20

## RESULT 4

AE2440  
hypothetical protein all5077 [imported] - Nostoc sp. (strain PCC 7120)  
C:Species: Nostoc sp. PCC 7120  
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002  
C:Accession: AE2440  
R:Kaneko, T.; Nakamura, Y.; Molk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriyuchi,  
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S  
DNA Res. 8, 205-213, 2001  
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana  
A:Reference number: AB1807; MUID:21395285; PMID:11759640  
A:Accession: AE2440  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-252 <KUR>  
A:Cross-references: GB:BA000019; PIDN:BAE76776.1; PID:g17134215; GSPDB:GN00179  
A:Experimental source: strain PCC 7120  
C:Genetics:  
A:Gene: all5077

Query Match 0.4%; Score 8; DB 2; Length 252;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1609 TNGFSLDS 1616  
|||  
Db 126 TNGFSLDS 133

## RESULT 5

F95887  
probable ABC transporter ATP-binding protein SMB20380 [imported] - Sinorhizobium meliloti  
C:Species: Sinorhizobium meliloti  
C:Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 19-Jul-2002  
C:Accession: F95887  
R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernanc  
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001  
A:Title: The complete sequence of the 1.683-kb PSYMB megaplasmid from the N2-fixing endo  
A:Reference number: A95842; MUID:21396508; PMID:11481431  
A:Accession: F95887  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-349 <KUR>  
A:Cross-references: GB:AL591985; PIDN:CAC48766.1; PID:g15140239; GSPDB:GN00167  
A:Experimental source: strain 1021, megaplasmid PSYMB  
R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abolia, P.; Ampe, F.; Barloy-Hubler,  
pela, D.; Chain, P.; Cowie, A.; Davies, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;  
Science 293, 668-672, 2001  
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,  
hehalt, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.  
A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.  
A:Reference number: A96039; MUID:21368234; PMID:11474104  
C:Contents: annotation  
C:Genetics:  
A:Gene: SMB20380  
A:Genome: plasmid  
C:Superfamily: inner membrane protein malK; ATP-binding cassette homology

Query Match 0.4%; Score 8; DB 2; Length 349;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 370 TYDGRHNA 377  
|||  
Db 14 TYDGRHNA 21

## RESULT 6

T04089  
GTP-binding protein beta chain (clone Gbeta2) - common tobacco  
C:Species: Nicotiana tabacum (common tobacco)

C>Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 20-Jun-2000  
 C/Accession: T04089  
 R/Lein, W.; Saalbach, G.  
 Submitted to the EMBL Data Library, February 1997  
 A/Description: Characterization of G-protein function in plants.  
 A/Reference number: Z15198  
 A/Accession: T04089  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: mRNA  
 A/Residues: 1-377 <LEI>  
 A/Cross-references: EMBL:Z84821; PIDD:CA80619.1  
 A/Experimental source: strain SRI; tissue-type leaf; clone Gbet2  
 A/Superfamily: GTP-binding regulatory protein beta chain; WD repeat homology  
 F/244-277/Domain: WD repeat homology <MDR>

Query Match 0.4%; Score 8; DB 2; Length 377;  
 Best Local Similarity 100.0%; Pred. No. 25;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 667 MSVKEKE 674  
 Db 1 MSVKEKE 8

RESULT 7  
 B71262  
 probable glutamate transporter - syphilis spirochete  
 C/Species: Treponema pallidum subsp. pallidum (syphilis spirochete)  
 C/Date: 24-Jul-1998 #sequence\_revision 24-Jul-1998 #text\_change 05-Nov-1999  
 C/Accession: B71262  
 R/Frazer, C.M.; Norris, S.J.; Weinstein, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin  
 reon, J.; Khajak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Uterback, T.; McD  
 they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.  
 Science 281, 375-388, 1998  
 A/Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.  
 A/Reference number: A71250; MUID:98332770; PMID:9665876  
 A/Accession: B71262  
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-407 <COL>  
 A/Cross-references: GB:AE001262; GB:AE000520; NID:G3323254; PIDD:MAC65891.1; PID:G332325  
 A/Experimental source: strain Nichols  
 C/Genetics:  
 A/Gene: TP0934

Query Match 0.4%; Score 8; DB 2; Length 407;  
 Best Local Similarity 100.0%; Pred. No. 26;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 268 RRELLLR 275  
 Db 61 RRELLLR 68

RESULT 8  
 BVBVMS  
 MS11 protein - yeast (Saccharomyces cerevisiae)  
 N/Alternate names: protein YBR1406; protein YBR195c  
 C/Species: Saccharomyces cerevisiae  
 C/Date: 30-Jun-1991 #sequence\_revision 30-Jun-1991 #text\_change 05-Nov-1999  
 C/Accession: S07865; S34020; S46067  
 R/Ruggieri, R.; Tanaka, K.; Nakafuku, M.; Kaziro, Y.; Toh-e, A.; Matsunoto, K.  
 Proc Natl Acad Sci U S A 86, 8778-8782, 1989  
 A/Title: MS11, a negative regulator of the RAS-CAMP pathway in Saccharomyces cerevisiae.  
 A/Reference number: S07865; MUID:90046875; PMID:2554329  
 A/Accession: S07865  
 A/Molecule type: DNA  
 A/Residues: 1-422 <RUG>  
 A/Cross-references: GB:M27300; NID:G172005; PIDD:AAA34804.1; PID:G172006  
 R/Demolis, N.; Maillet, L.; Bussereau, F.; Jacquet, M.  
 Yeast 9, 645-659, 1993  
 A/Title: RIM2, MS11 and PGI1 are located within an 8 kb segment of Saccharomyces cerevis  
 leucine zipper motif.

A/Reference number: S33966; MUID:93348777; PMID:8346681  
 A/Accession: S34020  
 A/Status: translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-422 <DEM>  
 A/Cross-references: EMBL:Z21487; NID:G311665; PIDD:CAA79682.1; PID:G311671  
 R/Bussereau, F.; Demolis, N.; Jacquet, M.; Maillet, L.  
 Submitted to the Protein Sequence Database, August 1994  
 A/Reference number: S46054  
 A/Accession: S46067  
 A/Molecule type: DNA  
 A/Residues: 1-422 <BUS>  
 A/Cross-references: EMBL:Z26064; NID:G536562; PIDD:CAA85157.1; PID:G536563; GSPDB:GN00002  
 C/Genetics:  
 A/Gene: SGD:MS11; MIPS:YBR195c  
 A/Cross-references: SGD:S0000399; MIPS:YBR195c  
 A/Map position: 2R  
 A/Superfamily: MS11 protein; WD repeat homology  
 C/Keywords: duplication  
 F/196-230/Domain: WD repeat homology <WD1>  
 F/292-326/Domain: WD repeat homology <WD2>  
 F/335-371/Domain: WD repeat homology <WD3>  
 F/380-413/Domain: WD repeat homology <WD4>

Query Match 0.4%; Score 8; DB 1; Length 422;  
 Best Local Similarity 100.0%; Pred. No. 27;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 99 TEGNAVSL 106  
 Db 271 TEGNAVSL 278

RESULT 9  
 D95194  
 hypothetical protein SPI670 [imported] - Streptococcus pneumoniae (strain TIGR4)  
 C/Species: Streptococcus pneumoniae  
 C/Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 03-Aug-2001  
 C/Accession: D95194  
 R/Fretzelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heide  
 on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple, E  
 nson, T.; Hickey, E.K.; Holt, I.E.  
 Science 293, 498-506, 2001  
 A/Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,  
 A/Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.  
 A/Reference number: A95000; MUID:21357209; PMID:11463916  
 A/Accession: D95194  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-457 <KUR>  
 A/Cross-references: GB:AE005672; PIDD:AAK75749.1; PID:G14973162; GSPDB:GN00164; TIGR:SP46  
 A/Experimental source: strain TIGR4  
 C/Genetics:  
 A/Gene: SPI670

Query Match 0.4%; Score 8; DB 2; Length 457;  
 Best Local Similarity 100.0%; Pred. No. 29;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 542 VSEEQIRL 549  
 Db 290 VSEEQIRL 297

RESULT 10  
 A98061  
 UDP-N-acetylglucosaminyl-D-glucamyl-lysine-D-alanyl-D-alanine ligase (EC 6.3.2.10) [im  
 C/Species: Streptococcus pneumoniae  
 C/Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 03-Jun-2002  
 C/Accession: A98061  
 R/Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E  
 e, R.; Leblanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M  
 y, P.; Sun, P.M.; Winkler, M.E.

J. Bacteriol. 183, 5709-5717, 2001  
 A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;  
 A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.  
 A:Reference number: A97872; MUID:21429245; PMID:11544234  
 A:Accession: A98061  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-457 <KUR>  
 A:Cross-references: GB:AE007317, PIDN:AL00318.1; PID:G15459176; GSPDB:GN00174  
 C:Genetics:  
 A:Gene: murF  
 C:Keywords: ligase

Query Match 0.4%; Score 8; DB 2; Length 457;  
 Best Local Similarity 100.0%; Pred. No. 29;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 542 VSEQIRL 549  
 |||||  
 Db 290 VSEQIRL 297

RESULT 11  
 T31747  
 hypothetical protein C05C8.7 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 21-Jan-2000  
 C:Accession: T31747  
 R:Sammons, L.; Woldmann, P.  
 submitted to the EMBL Data Library, July 1997  
 A:Description: The sequence of C. elegans coamid C05C8.  
 A:Reference number: Z21078  
 A:Accession: T31747  
 A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA  
 A:Residues: 1-467 <SAM>  
 A:Cross-references: EMBL:AF016430; PIDN:AAB65374.1; GSPDB:GN00023; CESP:C05C8.7  
 C:Genetics:  
 A:Experimental source: strain Bristol N2; clone C05C8

A:Gene: CESP:C05C8.7  
 A:Map position: 5  
 A:introns: 43/3; 106/3; 154/1; 214/3; 411/2  
 C:Superfamily: yeast mannose-6-phosphate isomerase

Query Match 0.4%; Score 8; DB 2; Length 467;  
 Best Local Similarity 100.0%; Pred. No. 29;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 117 GLRGAVVE 124  
 |||||  
 Db 227 GLRGAVVE 234

RESULT 12  
 S62566  
 PHD finger protein - fission yeast (Schizosaccharomyces pombe)  
 C:Species: Schizosaccharomyces pombe  
 C:Date: 16-May-1996 #sequence\_revision 13-Mar-1997 #text\_change 21-Jan-2000  
 C:Accession: T38591; S62566  
 R:Pearson, D.; Churchill, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.  
 submitted to the EMBL Data Library, November 1995  
 A:Reference number: Z21801  
 A:Accession: T38591  
 A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA  
 A:Residues: 1-538 <PR2>  
 A:Cross-references: EMBL:Z67961; NID:G1065887; PIDN:CAA91894.1; PID:G1065895; GSPDB:GN00  
 A:Experimental source: strain 972h-; cosmid G30D11  
 C:Genetics:  
 A:Gene: SPAC30D11.08c  
 A:Map position: 1L  
 A:introns: 447/2

Query Match 0.4%; Score 8; DB 2; Length 538;  
 Best Local Similarity 100.0%; Pred. No. 33;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 237 EESNONGG 244  
 |||||  
 Db 41 EESNONGG 48

RESULT 13

E86177  
 hypothetical protein [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
 C:Accession: E86177

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
 ansen, N.F.; Hughes, B.; Huitzar, L.  
 Nature 408, 816-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luo, J.S.; Maiti, R.; Marziani,  
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shin, P.; Southwick, A.M.; Sun, H.; Tallon,  
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A:Reference number: A86141; MUID:21016719; PMID:11130712  
 A:Accession: E86177  
 A:Status: preliminary

A:Molecule type: DNA  
 A:Residues: 1-580 <STO>  
 A:Cross-references: GB:AE005172; NID:G2341025; PIDN:AAB70423.1; GSPDB:GN00141  
 C:Genetics:  
 A:Map position: 1

Query Match 0.4%; Score 8; DB 2; Length 580;  
 Best Local Similarity 100.0%; Pred. No. 35;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 452 TDLVLTAS 459  
 |||||  
 Db 359 TDLVLTAS 366

RESULT 14

JC4642  
 purH bifunctional enzyme - human

N:Alternate names: 5-aminoimidazole-4-carboxamide ribonucleotide transformylase  
 N:Contains: IMP cyclohydrolase (EC 3.5.4.10); phosphoribosylaminoimidazolecarboxamide fo  
 C:Species: Homo sapiens (man)  
 C:Date: 10-May-1996 #sequence\_revision 19-Jul-1996 #text\_change 20-Jun-2000  
 C:Accession: JC4642; JCS578

R:Yamauchi, M.; Seki, N.; Mita, K.; Saito, T.; Tsuji, S.; Hongo, E.; Moriyo, M.; Shiomi  
 DNA Res. 2, 269-275, 1995  
 A:Title: Isolation of human purH gene expressed in the rodent transformant cells by sub  
 A:Reference number: JC4642; MUID:97021441; PMID:8867801  
 A:Accession: JC4642

A:Molecule type: mRNA  
 A:Residues: 1-592 <YAM>  
 A:Cross-references: DDBJ:DB2348; NID:G1311461; PIDN:BA11559.1; PID:G1311462  
 A:Experimental source: testis  
 R:Segita, T.; Aya, H.; Ueno, M.; Ishizuka, T.; Kawashima, K.  
 J. Biochem. 122, 309-313, 1997

A:Title: Characterization of molecularly cloned human 5-aminoimidazole-4-carboxamide rib  
 A:Reference number: JCS578; MUID:98021067; PMID:9378707

A:Accession: JCS578  
 A:Molecule type: DNA  
 A:Residues: 1-592 <STUG>  
 A:Cross-references: DDBJ:D89976; NID:G2317691; PIDN:BA421762.1; PID:G2317692

A:Experimental source: placenta cell  
 C:Comment: This enzyme catalyzes formylation of 5-aminoimidazole-4-carboxamide ribonucle  
 C:Genetics:  
 A:Gene: purH  
 C:Superfamily: purH bifunctional enzyme



C;Keywords: hydrolase; testis; transferase  
F;470-481/Domain: N10-formylatedtrihydrofolate binding #status predicted <NFB>  
F;213,267/Active site: His #status predicted

Query Match 0.4%; Score 8; DB 2; Length 592;  
Best Local Similarity 100.0%; Pred. No. 36;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 669 VKELKEAL 676  
|||  
Db 250 VKELKEAL 257

RESULT 15

A28088

oxaloacetate decarboxylase (EC 4.1.1.3) alpha chain - Klebsiella pneumoniae

N;Alternate names: oxalate beta-decarboxylase, alpha chain

C;Species: Klebsiella pneumoniae

C;Date: 19-May-1989 #sequence\_revision 19-May-1989 #text\_change 11-Jan-2002

C;Accession: A28088

R;Schwarz, E.; Oesterhelt, D.; Reinke, H.; Beyreuther, K.; Dimroth, P.

J. Biol. Chem. 263, 9640-9645, 1988

A;Title: The sodium ion translocating oxalacetate decarboxylase of Klebsiella pneumoniae

A;Reference number: A28088; MUID:88237085; PMID:2454915

A;Accession: A28088

A;Molecule type: DNA

A;Residues: 1-596 <SCH>

A;Cross-references: EMBL:J03885; NID:G149288; PID:AAA25120.1; PID:G149289

C;Superfamily: Klebsiella pneumoniae oxaloacetate decarboxylase alpha chain; lipoyl/biot

C;Keywords: biotin binding; carbon-carbon lyase; carboxy-lyase; sodium pump

F;553-596/Domain: lipoyl/biotin-binding homology <LPB>

F;562/Binding site: biotin (Lys) (covalent) #status predicted

Query Match 0.4%; Score 8; DB 2; Length 596;  
Best Local Similarity 100.0%; Pred. No. 36;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 277 EVLAIPR 284  
|||  
Db 318 EVLAIPR 325

Search completed: January 2, 2004, 16:15:29  
Job time : 32 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2004, 16:07:03 ; Search time 21 Seconds

(without alignments)  
4010.706 Million cell updates/sec

Title: US-09-983-025a-2

Sequence: 1 MMCKILRLISAILAGWALC.....AADCDLDECTCRDPKAEENQ 1791

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 127863 seqs, 47026705 residues

Word size : 0

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	18	1.0	1627	1	PAPA_HUMAN
2	8	0.4	105	1	RL24_THEMA
3	8	0.4	377	1	GBB2_TOBAC
4	8	0.4	422	1	MS11_YEAST
5	8	0.4	533	1	SBR4_RAT
6	8	0.4	538	1	YAU8_SCHPO
7	8	0.4	592	1	PUR9_HUMAN
8	8	0.4	595	1	PCOA_KLEBN
9	8	0.4	4563	1	APB_HUMAN
10	7	0.4	103	1	RL24_BACHD
11	7	0.4	103	1	RL24_BACST
12	7	0.4	103	1	RL24_BACSU
13	7	0.4	103	1	RL24_ECOLI
14	7	0.4	103	1	RL24_HAEIN
15	7	0.4	104	1	RL24_BUCAN
16	7	0.4	113	1	RL24_MICLU
17	7	0.4	115	1	RL24_SYNY3
18	7	0.4	132	1	HEX9_ADS40
19	7	0.4	147	1	FER1_SPILO
20	7	0.4	149	1	VLI_HPV62
21	7	0.4	154	1	RISB_AQUAE
22	7	0.4	155	1	BCTI_BOVIN
23	7	0.4	155	1	BCTI_SHEEP
24	7	0.4	161	1	CC31_YEAST
25	7	0.4	169	1	RIMM_NEIMA
26	7	0.4	169	1	RIMM_NEIMA
27	7	0.4	196	1	YRAO_ECOLI
28	7	0.4	205	1	KITH_ECOLI
29	7	0.4	205	1	KITH_ECOLI
30	7	0.4	236	1	VHEL_WCMVM
31	7	0.4	236	1	VHEL_WCMVO
32	7	0.4	237	1	NANE_THETN
33	7	0.4	244	1	YENR_YEREN

34	7	0.4	250	1	PSA4_DICDI
35	7	0.4	268	1	Y237_THEAC
36	7	0.4	270	1	REP8_HUMAN
37	7	0.4	279	1	YERC_SCHPO
38	7	0.4	314	1	YFDV_ECOLI
39	7	0.4	322	1	PRIM_ECOLI
40	7	0.4	325	1	PE68_ARATH
41	7	0.4	334	1	SELD_MERKA
42	7	0.4	336	1	TOIB_MOUSE
43	7	0.4	336	1	Y4JT_RHISN
44	7	0.4	352	1	P53_ORYLA
45	7	0.4	354	1	ADA_ECOLI

## ALIGNMENTS

RESULT 1

PAPA\_HUMAN STANDARD: PRT; 1627 AA.

AC Q13219: Q08371: Q9UDK7:

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Pregnancy-associated plasma protein-A precursor (BC 3.4.24.-) (PAPP-A)

DE (Insulin-like growth factor-dependent IGF binding protein-4 protease)

DE (IGF-dependent IGFBP-4 protease) (IGFBP-4ase).

GN PAPA.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI\_TaxId:9606;

[1]

SEQUENCE FROM N.A., AND INDUCTION.

RC TISSUE=Placenta;

RX MEDLINE=96203921; PubMed=8620868;

RA Haaling J., Oxvig C., Overgaard M.T., Ebbeesen P., Kristensen T.,

RA Sottrup-Jensen L.;

RT "Complete cDNA sequence of the preproform of human pregnancy-associated plasma protein-A. Evidence for expression in the brain and induction by cAMP."

RT Eur. J. Biochem. 237:159-163(1996).

[2]

SEQUENCE OF 77-1627 FROM N.A., SEQUENCE OF 81-98; 117-126; 210-224; 466-485; 507-519; 576-593; 609-621; 718-736; 742-754; 1006-1017; 1289-1273; 1369-1374; 1389-1398; 1490-1509; 1524-1533 AND 1537-1544, VARIANT SER-944, AND TISSUE SPECIFICITY.

RC TISSUE=Placenta, and Serum;

RX MEDLINE=94146014; PubMed=7508748;

RA Kristensen T., Oxvig C., Sand O., Moller N.P.H., Sottrup-Jensen L.;

RA "Amino acid sequence of human pregnancy-associated plasma protein-A derived from cloned cDNA."

RT Biochemistry 33:1592-1598(1994).

[3]

SEQUENCE OF 81-99; 117-126; 210-224; 460-485; 507-519; 576-593; 718-736; 742-754; 1259-1273; 1369-1374; 1490-1509; 1524-1533 AND 1537-1544, SUBUNITS, AND INTERCHAIN DISULFIDE BOND.

RC TISSUE=Serum;

RX MEDLINE=93286045; PubMed=7685339;

RA Oxvig C., Sand O., Kristensen T., Gleich G.J., Sottrup-Jensen L.;

RA "Circulating human pregnancy-associated plasma protein-A is disulfide-bridged to the proform of eosinophil major basic protein."

RT J. Biol. Chem. 268:12243-12246(1993).

[4]

PARTIAL SEQUENCE, CARBOHYDRATE-LINKAGE SITES, AND DISULFIDE BONDS.

RC MEDLINE=22421368; PubMed=12421832;

RA Overgaard M.T., Sorensen E.S., Stachowiak D., Boldt H.B.,

RA Kristensen L., Sottrup-Jensen L., Oxvig C.;

RA "Complex of pregnancy-associated plasma protein-A and the proform of eosinophil major basic protein. Disulfide structure and carbohydrate attachment sites."

RT J. Biol. Chem. 278:2106-2117(2003).

[5]

RP IDENTIFICATION, FUNCTION, SUBCELLULAR LOCATION, AND TISSUE  
 RP SPECIFICITY.  
 RC TISSUE=Fibroblast;  
 RX MEDLINE=99179030; PubMed=10077652;  
 RA Lawrence J.B., Oxvig C., Overgaard M.T., Sottrup-Jensen L.,  
 RA Gleich G.J., Hays L.G., Yates J.R. III, Conover C.A.;  
 RT "The insulin-like growth factor (IGF)-dependent IGF binding protein-4  
 RT protease secreted by human fibroblasts is pregnancy-associated plasma  
 RT protein-A";  
 RL Proc. Natl. Acad. Sci. U.S.A. 96:3149-3153(1999).  
 RL [6]  
 RN FUNCTION, SUBUNITS, AND ENZYME REGULATION.  
 RP MEDLINE=20469470; PubMed=10913121;  
 RX Overgaard M.T., Haaning J., Boldt H.B., Olsen I.M., Laursen L.S.,  
 RA Christensen M., Gleich G.J., Sottrup-Jensen L., Conover C.A.,  
 RA Oxvig C.;  
 RT "Expression of recombinant human pregnancy-associated plasma protein-A  
 RT and identification of the proform of eosinophil major basic protein  
 RT as its physiological inhibitor";  
 RL J. Biol. Chem. 275:31128-31133(2000).  
 RL [7]  
 RN TISSUE SPECIFICITY.  
 RP MEDLINE=95057018; PubMed=7526035;  
 RX Bonno M., Oxvig C., Kephart G.M., Wagner J.M., Kristensen T.,  
 RA Sottrup-Jensen L., Gleich G.J.;  
 RT "Localization of pregnancy-associated plasma protein-A and  
 RT ribonucleic acid and eosinophil granule major basic protein messenger  
 RT ribonucleic acid in placenta";  
 RL Lab. Invest. 71:560-566(1994).  
 RL [8]  
 RN TISSUE SPECIFICITY, AND DEVELOPMENTAL STAGE.  
 RP MEDLINE=99423540; PubMed=10491647;  
 RX Overgaard M.T., Oxvig C., Christensen M., Lawrence J.B.,  
 RA Conover C.A., Gleich G.J., Sottrup-Jensen L., Haaning J.;  
 RT "Messenger ribonucleic acid levels of pregnancy-associated plasma  
 RT protein-A and the proform of eosinophil major basic protein:  
 RT expression in human reproductive and nonreproductive tissues";  
 RL Biol. Reprod. 61:1083-1089(1999).  
 RL [9]  
 RN DEVELOPMENTAL STAGE.  
 RP MEDLINE=95293954; PubMed=7539791;  
 RX Oxvig C., Haaning J., Kristensen L., Wagner J.M., Rubin I.,  
 RA Strigbrand T., Gleich G.J., Sottrup-Jensen L.;  
 RT "Identification of angiotensinogen and complement C3dg as novel  
 RT proteins binding the proform of eosinophil major basic protein in  
 RT human pregnancy serum and plasma";  
 RL J. Biol. Chem. 270:13645-13651(1995).  
 CC -1- FUNCTION: Metalloproteinase which specifically cleaves IGFBP-4 in  
 CC the presence of IGF, resulting in release of bound IGF.  
 CC -1- COFACTOR: Binds 1 zinc ion per subunit (By similarity).  
 CC -1- ENZYME REGULATION: Inhibited by complexation with the proform  
 CC of PRG2.  
 CC -1- SUBUNIT: Homodimer; disulfide-linked. In pregnancy serum,  
 CC predominantly found as a disulfide-linked 2:2 heterotrimer with  
 CC the proform of PRG2.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- TISSUE SPECIFICITY: High levels in placenta and pregnancy serum.  
 CC and in syncytiotrophoblasts in the chorionic villi. Lower levels  
 CC are found in a variety of other tissues including kidney,  
 CC myometrium, endometrium, ovaries, breast, prostate, bone marrow,  
 CC colon, fibroblasts and osteoblasts.  
 CC -1- DEVELOPMENTAL STAGE: Present in serum and placenta during  
 CC pregnancy; levels increase throughout pregnancy.  
 CC -1- INDUCTION: By 8-bromadenosine-3',5'-phosphate.  
 CC -1- PTM: There appear to be no free sulhydryl groups.  
 CC -1- SIMILARITY: Contains 5 Sushi (SCR) domains.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M46.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC -----  
 CC EMBL; U28727; AAC50543.1; -;  
 CC EMBL; X68280; CAA48341.1; -;  
 CC PIR; S65464; S65464.  
 CC MEROPS; M46.001; -;  
 CC Genew; HGNC:8602; PAPPA.  
 CC MIM; 176385; -;  
 CC GO; GO:0008237; F:metallopeptidase activity; IDA.  
 CC GO; GO:0008270; F:zinc ion binding activity; NAS.  
 CC GO; GO:0007565; P:pregnancy; NAS.  
 CC InterPro; IPR006558; Lamg-like.  
 CC InterPro; IPR000800; Notch.  
 CC InterPro; IPR000436; Sushi\_SCR\_CCP.  
 CC InterPro; IPR006025; Zn\_MTPeptidase.  
 CC Pfam; PF00084; sushi; 4.  
 CC SMART; SM00032; CCP; 4.  
 CC SMART; SM00560; LamGL; 1.  
 CC SMART; SM00004; NL; 3.  
 CC PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 CC Hydrolase; Metalloproteinase; Metal-binding; Zinc; Signal; Glycoprotein;  
 CC Zymogen; Repeat; Sushi.  
 CC SIGNAL  
 CC FT PROPEP 23 80  
 CC FT CHAIN 81 1627  
 CC FT DOMAIN 24 83  
 CC FT DOMAIN 272 583  
 CC FT DOMAIN 1215 1280  
 CC FT DOMAIN 1285 1342  
 CC FT DOMAIN 1346 1410  
 CC FT DOMAIN 1415 1471  
 CC FT DOMAIN 1478 1554  
 CC FT METAL 562 562  
 CC FT ACT SITE 563 563  
 CC FT METAL 566 566  
 CC FT DISULFID 144 235  
 CC FT DISULFID 327 622  
 CC FT DISULFID 332 657  
 CC FT DISULFID 414 428  
 CC FT DISULFID 424 440  
 CC FT DISULFID 457 473  
 CC FT DISULFID 461 461  
 CC FT DISULFID 474 485  
 CC FT DISULFID 583 600  
 CC FT DISULFID 587 612  
 CC FT DISULFID 710 878  
 CC FT DISULFID 713 881  
 CC FT DISULFID 732 732  
 CC FT DISULFID 753 835  
 CC FT DISULFID 775 781  
 CC FT DISULFID 947 975  
 CC FT DISULFID 960 971  
 CC FT DISULFID 983 990  
 CC FT DISULFID 999 1011  
 CC FT DISULFID 1036 1070  
 CC FT DISULFID 1051 1139  
 CC FT DISULFID 1192 1205  
 CC FT DISULFID 1210 1210  
 CC FT DISULFID 1215 1269  
 CC FT DISULFID 1227 1238  
 CC FT DISULFID 1242 1280  
 CC FT DISULFID 1285 1329  
 CC FT DISULFID 1300 1310  
 CC FT DISULFID 1314 1342  
 CC FT DISULFID 1346 1399  
 CC FT DISULFID 1362 1373  
 CC FT DISULFID 1377 1410  
 CC FT DISULFID 1415 1458  
 CC FT DISULFID 1428 1438  
 CC FT DISULFID 1442 1471  
 CC  
 CC INTERCHAIN (WITH C-51 OF PRG2 PROFORM).  
 CC OR 583-612.  
 CC OR 587-600.  
 CC  
 CC INTERCHAIN (WITH C-169 OF PRG2 PROFORM).  
 CC  
 CC INTERCHAIN.  
 CC

Query Match 1.0%; Score 18; DB 1; Length 1627;  
Best Local Similarity 100.0%; Pred. No. 1.9e-09;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 818 FTNQVARMHCYLDLVYQ 835  
DB 647 FTNQVARMHCYLDLVYQ 664

## RESULT 2

RL24\_THEME STANDARD; PRT; 105 AA.  
AC P38513;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE 50S ribosomal protein L24.  
GN RPLX OR TM1489.  
OS Thermotoga maritima.  
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.  
OX NCBI\_TaxID=2336;  
RN [1]  
RP STRAIN=MSB8 / DSM 3109;  
RC MEDLINE=95095941; PubMed=8002596;  
RA Sannangelantoni A.M., Bocchetta M., Cammarano P., Tiboni O.;  
RT "Phylogenetic depth of S10 and spc operons: cloning and sequencing of  
a ribosomal protein gene cluster from the extremely thermophilic  
RT bacterium Thermotoga maritima."  
RL J. Bacteriol. 176:7703-7710(1994).  
RN [2]  
RP STRAIN=MSB8 / DSM 3109;  
RC MEDLINE=99287316; PubMed=10360571;  
RA Nelson K.E., Clayton R.A., Gill S.R., Gwin M.L., Dodson R.J.,  
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,  
RA McDonald L., Ueberback T.R., Malek J.A., Linher K.D., Garrett M.M.,  
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,  
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eichen J.A., White O.,  
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;  
RT "Evidence for lateral gene transfer between Archaea and Bacteria from  
RT genome sequence of Thermotoga maritima."  
RL Nature 399:323-329(1999).  
CC -1- FUNCTION: THIS PROTEIN IS FOUND IN THE RIBONUCLEOPROTEIN CORE AND  
CC IS INVOLVED IN THE EARLY ASSEMBLY OF THE 50S SUBUNIT. IT IS NOT  
CC INVOLVED IN THE FUNCTIONS OF THE MATURE 50S SUBUNIT  
CC (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE L24P FAMILY OF RIBOSOMAL PROTEINS.  
CC  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; Z21677; CAAT9788.1; -  
CC EMBL; AE001798; AAD36555.1; -  
CC PIR; S40199; S40199.  
CC TIGR; TM1489; -  
CC InterPro; IPR005824; KOW.  
CC InterPro; IPR006646; KOW\_sub.  
CC InterPro; IPR003256; Ribosomal\_L24.  
CC InterPro; IPR005825; Ribosomal\_L24\_26.  
CC Pfam; PF00467; KOW; 1.  
CC ProDom; PD001677; Ribosomal\_L24; 1.  
CC SMART; SMO0739; KOW; 1.  
CC TIGRPFAMs; TIGR01079; rplX\_bact; 1.  
CC PROSITE; PS01109; RIBOSOMAL\_L24; 1.  
CC Ribosomal protein; Complete proteome.  
CC SEQUENCE 105 AA; 12023 MW; E7BCCP1B553C43D8 CRC64;

Query Match 0.4%; Score 8; DB 1; Length 105;  
Best Local Similarity 100.0%; Pred. No. 4.3;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 327 SGKDKGR 334  
DB 13 SGKDKGR 20

## RESULT 3

GBB2\_TOBAC STANDARD; PRT; 377 AA.  
AC P93358;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Guanine nucleotide-binding protein beta subunit 2.  
OS Nicotiana tabacum (Common tobacco).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Asteridae; Lamiales; Solanales; Solanaceae; Nicotiana.  
OX NCBI\_TaxID=4097;  
RN [1]  
RP STRAIN=cv. SRL; TISSUE=Leaf;  
RC Lein W., Saalbach G.;  
RA Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.  
RL  
CC -1- FUNCTION: GUANINE NUCLEOTIDE-BINDING PROTEIN (G PROTEINS) ARE  
CC INVOLVED AS A MODULATOR OR TRANSDUCER IN VARIOUS TRANSMEMBRANE  
CC SIGNALING SYSTEMS. THE BETA AND GAMMA CHAINS ARE REQUIRED FOR THE  
CC GTPASE ACTIVITY, FOR REPLACEMENT OF GDP BY GTP, AND FOR G PROTEIN-  
CC EFFECTOR INTERACTION.  
CC  
CC -1- SUBUNIT: G proteins are composed of 3 units (alpha, beta and  
CC gamma).  
CC  
CC -1- SIMILARITY: Contains 7 WD repeats.  
CC  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; Z84821; CAB06619.1; -  
CC PIR; T04089; T04089.  
CC HSSP; P04901; 1TBG.  
CC InterPro; IPR001632; Gproteins\_B.  
CC InterPro; IPR001680; WD40.  
CC Pfam; PF00400; WD40; 7.  
CC PRINTS; PR00319; GPROTEINB.  
CC PRINTS; PR00320; GPROTEINBPT.  
CC ProDom; PD0000018; WD40; 4.  
CC SMART; SMO0320; WD40; 7.  
CC PROSITE; PSS00678; WD\_REPEATS\_1; 2.  
CC PROSITE; PSS00678; WD\_REPEATS\_2; 5.  
CC PROSITE; PSS0294; WD\_REPEATS\_REGION; 1.  
CC Transducer; Repeat; WD repeat; Multigene family.  
CC  
CC REPEAT 63 93 WD 1.  
CC REPEAT 105 135 WD 2.  
CC REPEAT 154 185 WD 3.  
CC REPEAT 202 233 WD 4.  
CC REPEAT 246 276 WD 5.  
CC REPEAT 293 323 WD 6.  
CC REPEAT 339 369 WD 7.  
CC SEQUENCE 377 AA; 41084 MW; 2EF2BBA4272D3472 CRC64;

Query Match 0.4%; Score 8; DB 1; Length 377;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 667 MSYKELKE 674

Db 1 MSVKEKE 8

```

RESULT 4
MS11 YEAST STANDARD; PRT; 422 AA.
ID MS11 YEAST
AC P13712;
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-JAN-1990 (Rel. 13, Last annotation update)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Chromatin assembly factor 1 P50 subunit (CAF-1 P50 subunit) (MS11
DE protein) (IRAI multicopy suppressor).
GN MS11 OR CAC3 OR YBR195C OR YBR1405.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90046875; PubMed=2554329;
RA Ruggieri R., Tanaka K., Nakafuku M., Kaziro Y., Toh-E A.,
RA Matsunoto K.;
RT "MS11, a negative regulator of the RAS-CAMP pathway in Saccharomyces
RT cerevisiae."
RL Proc. Natl. Acad. Sci. U.S.A. 86:8778-8782(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=5288C;
RX MEDLINE=93348777; PubMed=8346681;
RA Demolis N., Mallet L., Bussereau F., Jacquet M.;
RT "RIM2, MS11 and PGI1 are located within an 8 kb segment of
RT Saccharomyces cerevisiae chromosome II, which also contains the
RT putative ribosomal gene L21 and a new putative essential gene with a
RT leucine zipper motif."
RL Yeast 9:645-659(1993).
RN [3]
RP SEQUENCE OF 7-27, 32-43 AND 414-422, AND CHARACTERIZATION.
RX MEDLINE=97182548; PubMed=9030687;
RA Kaufman P.D., Kobayashi R., Stillman B.;
RT "Ultra-violet radiation sensitivity and reduction of telomeric
RT silencing in Saccharomyces cerevisiae cells lacking chromatin
RT assembly factor-1."
RL Genes Dev. 11:345-357(1997).
RN [4]
RP FUNCTION: COMPLEX THAT ASSEMBLES HISTONE OCTAMERS ONTO REPLICATING
RN DNA IN VITRO. CAF-1 PERFORMS THE FIRST STEP OF THE NUCLEOSOME
RN ASSEMBLY PROCESS, BRINGING NEWLY SYNTHESIZED HISTONES H3 AND H4 TO
RN REPLICATING DNA; HISTONES H2A/H2B CAN BIND TO THIS CHROMATIN
RN PRECURSOR SUBSEQUENT TO DNA REPLICATION TO COMPLETE THE HISTONE
RN OCTAMER. P150 AND P60 FORM COMPLEXES WITH NEWLY SYNTHESIZED
RN HISTONES H3 AND ACETYLATED H4 IN CELL EXTRACTS (BY SIMILARITY).
CC -1 SUBUNIT: COMPOSED OF THREE SUBUNITS, P50, P60 AND P90.
CC -1 SUBCELLULAR LOCATION: Nuclear.
CC -1 SIMILARITY: Contains 6 WD repeats.
CC -1 SIMILARITY: BELONGS TO THE RBAF46/RBAF48/MS11 FAMILY OF WD-REPEAT
CC PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M27300; AAA34804.1; -
CC EMBL; Z21487; CAA79682.1; -
CC EMBL; Z36064; CAA85157.1; -
CC PIR; S07865; BVBYSM.
CC SGP; S0000399; MS11.
CC GO; GO:0005678; C:chromatin assembly complex; IDA.
CC GO; GO:0005677; C:chromatin silencing complex; IMP.
CC GO; GO:0006342; P:chromatin silencing; IMP.

```

```

DR GO; GO:0006281; P:DNA repair; IMP.
DR GO; GO:0006334; P:nucleosome assembly; IMP.
DR InterPro: IPR001680; WD40.
DR Pfam; PF00400; WD40; 5.
DR SMART; SM00320; WD40; 5.
DR PROSITE; PS00678; WD_REPEATS_1; 3.
DR PROSITE; PS50082; WD_REPEATS_2; 2.
DR PROSITE; PS50294; WD_REPEATS_REGION; 1.
KW CAMP; Repeat; WD repeat.
FT REPEAT 127 158 WD 1.
FT REPEAT 198 238 WD 2.
FT REPEAT 249 289 WD 3.
FT REPEAT 294 334 WD 4.
FT REPEAT 338 379 WD 5.
FT REPEAT 382 421 WD 6.
SQ SEQUENCE 422 AA; 47364 MW; 0D3DB6CB2AC74166 CRC64;

Query Match 0.4%; Score 8; DB 1; Length 422;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 99 TEGNAVS 106
Db 271 TEGNAVS 278

RESULT 5
SERA RAT STANDARD; PRT; 533 AA.
ID SERA RAT
AC O08651;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE P-3-phosphoglycerate dehydrogenase (EC 1.1.1.95) (3-PGDH).
GN PGDH.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97250633; PubMed=9163325;
RA Achouri Y., Rider M.H., van Schaftingen E., Robbi M.;
RT "Cloning, sequencing and expression of rat liver 3-phosphoglycerate
RT dehydrogenase."
RL Biochem. J. 323:365-370(1997).
RN [2]
RP CATALYTIC ACTIVITY: 3-phosphoglycerate + NAD(+) = 3-
RP phosphohydroxypropylate + NADH.
CC -1 PATHWAY: Serine biosynthesis; first step.
CC -1 SUBUNIT: Homotrimer.
CC -1 TISSUE SPECIFICITY: LIVER, KIDNEY, BRAIN, TESTIS.
CC -1 SIMILARITY: BELONGS TO THE D-ISOMER SPECIFIC 2-HYDROXYACID
CC DEHYDROGENASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X97772; CAA6374.1; -
CC HSP; P08328; 1PSD.
CC InterPro: IPR006139; 2-Hacid_DH.
CC InterPro: IPR006140; 2-Hacid_DH_C.
CC InterPro: IPR006236; PGDH.
CC Pfam; PF00389; 2-Hacid_DH; 1.
CC Pfam; PF02826; 2-Hacid_DH_C; 1.
CC TIGRfam; TIGR01327; PGDH; 1.
CC PROSITE; PS00665; D_2_HYDROXYACID_DH_1; 1.
CC PROSITE; PS00670; D_2_HYDROXYACID_DH_2; 1.
CC PROSITE; PS00671; D_2_HYDROXYACID_DH_3; 1.

```

KW Serine biosynthesis; Oxidoreductase; NAD.  
 FT ACT\_SITE 236 236 SUBSTRATE BINDING (BY SIMILARITY).  
 FT ACT\_SITE 265 265 BY SIMILARITY.  
 FT ACT\_SITE 283 283 BY SIMILARITY.  
 SQ SEQUENCE 533 AA; 56493 MW; 7273DAC3349595EF CRC64;

Query Match 0.4%; Score 8; DB 1; Length 533;  
 Best Local Similarity 100.0%; Pred. No. 17;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1685 PSDPMLP 1692  
 DB 472 PSDPMLP 479

RESULT 6  
 YAB8\_SCHPO STANDARD; PRT; 538 AA.

AC Q009908;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Hypothetical protein C30D11.08c in chromosome I.  
 GN SPAC30D11.08c.  
 OS Schizosaccharomyces pombe (Fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomycetes.  
 NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RX MEDLINE=21844401; PubMed=11859360;  
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsis K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Moule S., Mungall K., Murphy L., Mablett D., Odell C.,  
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
 RA Taylor J., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Voiclaert G., Aert R., Robben J., Grymptre B.,  
 RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,  
 RA Borzym K., Langer I., Beck A., Lebrach H., Reinhardt R., Pohl T.M.,  
 RA Eger P., Zimmermann W., Medler H., Wambutt R., Purnelle B.,  
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mortier S.,  
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 RA Lucas M., Roche M., Galliard C., Tallada V.A., Garzon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Cerutti L., Lowe T., McCombie M.R., Paulsen I., Potashkin J.,  
 RA Snpakovski G.V., Uesery D., Barrell B.G., Nurse P.;  
 RT "The genome sequence of Schizosaccharomyces pombe.";  
 RL Nature 415:871-880(2002).  
 CC -1- SIMILARITY: TO S.POMBE SPCC43.07C AND SOME, TO C.ELEGANS ZK783.4.  
 CC -1- SIMILARITY: Contains 1 PHD-type zinc finger.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).

DR InterPro; IPR001965; Znf\_PHD.  
 DR Pfam; PF00628; PHD\_1.  
 DR SMART; SM00249; PHD; 1.  
 DR PROSITE; PS01359; ZF\_PHD\_1; 1.  
 DR PROSITE; PS50016; ZF\_PHD\_2; 1.  
 KW Hypothetical protein; Zinc-finger.  
 FT ZN\_FING 232 288 PHD-TYPE.  
 SQ SEQUENCE 538 AA; 60696 MW; 2FD754D50145982C CRC64;

Query Match 0.4%; Score 8; DB 1; Length 538;  
 Best Local Similarity 100.0%; Pred. No. 17;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 237 EESNONG 244  
 DB 41 EESNONG 48

RESULT 7  
 PUR9\_HUMAN STANDARD; PRT; 592 AA.

AC P31939; O13856;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Bifunctional purine biosynthesis protein PURH [Includes:  
 DE Phosphoribosylaminoimidazolecarboxamide formyltransferase (EC 2.1.2.3)  
 DE (AICAR transformylase); IMP cyclohydrolase (EC 3.5.4.10) (inosinase)  
 DE (IMP synthetase) (ATTC)].  
 GN ATIC OR PURH.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Hepatoma;  
 RX MEDLINE=96147205; PubMed=8567683;  
 RA Rayl E.A., Morison B.A., Beardsley G.P.;  
 RT "The human purh gene product, 5-aminomidazole-4-carboxamide  
 RT ribonucleotide formyltransferase/IMP cyclohydrolase. Cloning,  
 RT sequencing, expression, purification, kinetic analysis, and domain  
 RT mapping.";  
 RL J. Biol. Chem. 271:2225-2233(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RX MEDLINE=97021441; PubMed=8867801;  
 RA Yamauchi M., Seki N., Mita K., Saito T., Tsuji S., Hongo E.,  
 RA Moritomo M., Shiohara T., Koyama H.;  
 RT "Isolation of human purh gene expressed in the rodent transformatant  
 RT cells by subtractive enrichment of 3'-untranslated region of human  
 RT transcript.";  
 RL DNA Res. 2:269-275(1995).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RX MEDLINE=98021067; PubMed=9378707;  
 RA Sugita T., Aya H., Ueno M., Ishizuka T., Kawashima K.;  
 RT "Characterization of molecularly cloned human 5-aminomidazole-4-  
 RT carboxamide ribonucleotide transferase.";  
 RL J. Biochem. 122:309-313(1997).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Pancreas;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.T., Wang J., Heien F.,  
 RA Diatchenko L., Marusik K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Udell T.B., Tothiyuki S., Carninci P., Prange C.,

```

RESULT 8
DCCA_KLEPN STANDARD; PRT; 595 AA.
ID DCCA_KLEPN STANDARD; PRT; 595 AA.
AC P1387;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-JUN-1993 (Rel. 26, Last annotation update)
DE Oxaloacetate decarboxylase alpha chain (EC 4.1.1.3).
GN OADA.
OS Klebsiella pneumoniae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Klebsiella.
OX NCBI_Taxid=573;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88257085; PubMed=2454915;
RA Schwarz E., Oesterhelt D., Reinke H., Beyreuther K., Dimroth P.;
RT "The sodium ion translocating oxalacetate decarboxylase of Klebsiella
RT pneumoniae. Sequence of the biotin-containing alpha-subunit and
RT relationship to other biotin-containing enzymes.";
RL J. Biol. Chem. 263:9640-9645(1988).
CC -1- FUNCTION: LYASE AND SODIUM TRANSPORTER.
CC -1- CATALYTIC ACTIVITY: Oxaloacetate = pyruvate + CO(2).
CC -1- COFACTOR: BIOTIN AND REQUIRES A SODIUM ION.
CC -1- SUBUNIT: COMPOSED OF THREE CHAINS (ALPHA, BETA, AND GAMMA).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, J03885; AAA25120.1; -.
DR PIR, A28088; A28088.
DR HSSP, P02905; 1A6X.
DR InterPro: IPR001882; Biotin attach.
DR InterPro: IPR000089; Biotin_1ipoy1.
DR InterPro: IPR000891; HMGL-like.
DR InterPro: IPR005776; OADA.
DR InterPro: IPR003379; PYC_OADA.
DR Pfam; PF00364; biotin_1ipoy1; 1.
DR Pfam; PF00682; HMGL-like; 1.
DR Pfam; PF02436; PYC_OADA; 1.
DR TIGRFAMs; TIGR01108; oada; 1.
DR PROSITE; PS00188; BIOTIN; 1.
DR Decarboxylase; Lyase; Sodium transport; Biotin.
KW INT MET
FT BINDING 561 561 BIOTIN (BY SIMILARITY).
FT FT
SQ SEQUENCE 595 AA; 63402 MW; AA744A95A6E9488C CRC64;

Query Match 0.4%; Score 8; DB 1; Length 595;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 277 EVLAEIPR 284
|||||
Db 317 EVLAEIPR 324

RESULT 9
APB_HUMAN STANDARD; PRT; 4563 AA.
ID APB_HUMAN STANDARD; PRT; 4563 AA.
AC P04114; O00502; Q13787;
DT 01-NOV-1986 (Rel. 03, Created)
DT 01-NOV-1986 (Rel. 03, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Apolipoprotein B-100 precursor (Apo B-100) [Contains: Apolipoprotein
DE B-48 (Apo B-48)].
GN APOB.
OS Homo sapiens (Human).

```



- OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87016385; PubMed=3763409;  
 RA Knott T.C., Wallis S.C., Powell L.M., Pease R.J., Lusis A.J.,  
 RA Blackhart B., McCarthy B.J., Mahley R.W., Levy-Wilson B., Scott J.;  
 RT "Complete cDNA and derived protein sequence of human apolipoprotein  
 B-100.";  
 RL Nucleic Acids Res. 14:7501-7503(1986).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=88003974; PubMed=3652907;  
 RA Ludwig E.H., Blackhart B.D., Pierotti V.R., Calati L., Fortier C.,  
 RA Knott T., Scott J., Mahley R.W., Levy-Wilson B., McCarthy B.J.;  
 RT "DNA sequence of the human apolipoprotein B gene.";  
 RL DNA 6:363-372(1987).  
 [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87008488; PubMed=3759943;  
 RA Chen S.-H., Yang C.-Y., Chen P.-F., Setzer D., Tanimura M., Li W.-H.,  
 RA Goto A.M., Jr., Chan L.,  
 RT "The complete cDNA and amino acid sequence of human apolipoprotein  
 B-100.";  
 RL J. Biol. Chem. 261:12918-12921(1986).  
 [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87041416; PubMed=3464946;  
 RA Law S.W., Grant S.M., Higuchi K., Hospattankar A.V., Lackner K.J.,  
 RA Lee N., Brewer H.B., Jr.;  
 RT "Human liver apolipoprotein B-100 cDNA: complete nucleic acid and  
 RT derived amino acid sequence.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 83:8142-8146(1986).  
 [5]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87161758; PubMed=3030729;  
 RA Cladaras C., Hadjopoulos-Cladaras M., Nolte R.T., Atkinson D.,  
 RA Zannis V.I.;  
 RT "The complete sequence and structural analysis of human  
 RT apolipoprotein B-100: relationship between apob-100 and apob-48  
 RT forms.";  
 RL EMBO J. 5:3495-3507(1986).  
 [6]  
 RP SEQUENCE OF 709-906 FROM N.A.  
 RX MEDLINE=85270450; PubMed=3860836;  
 RA Deed S.S., Koculsky A.G., Albers J.U.;  
 RT "A partial cDNA clone for human apolipoprotein B.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 82:4983-4986(1985).  
 [7]  
 RP SEQUENCE OF 3056-3159 FROM N.A.  
 RX MEDLINE=86041888; PubMed=3903660;  
 RA Merridan M., Schumaker V.N., Fareed G.C., West R., Johnson D.F.,  
 RA Kirchgessner T.G., Lin H.-C., Wang X., Ma Y., Mendiaz E., Lusis A.J.;  
 RT "Human apolipoprotein B: identification of cDNA clones and  
 RT characterization of mRNA.";  
 RL Nucleic Acids Res. 13:6937-6953(1985).  
 [8]  
 RP SEQUENCE OF 1937-2018 AND 3811-4334 FROM N.A.  
 RX MEDLINE=86093680; PubMed=3841204;  
 RA Carlsson P., Olafsson S.O., Bondjers G., Darnfors C., Wiklund O.,  
 RA Bjursell G.;  
 RT "Molecular cloning of human apolipoprotein B cDNA.";  
 RL Nucleic Acids Res. 13:8813-8826(1985).  
 [9]  
 RP SEQUENCE OF 3109-4563 FROM N.A.  
 RX MEDLINE=85300528; PubMed=2994225;  
 RA Knott T.J., Rall S.C. Jr., Innerarity T.L., Jacobson S.F.,  
 RA Ureia M.S., Levy-Wilson B., Powell L.M., Pease R.J., Eddy R.,  
 RA Nakai H., Byers M., Priestley L.M., Robertson E., Rall L.B.,  
 RA Bethesda C., Shows T.B., Mahley R.W., Scott J.;  
 RT "Human apolipoprotein B: structure of carboxyl-terminal domains,  
 RT sites of gene expression, and chromosomal localization.";  
 RL Science 230:37-43(1985).  
 [10]  
 RP SEQUENCE OF 1-291 FROM N.A.  
 RX MEDLINE=86149325; PubMed=351177;  
 RA Procter A.A., Hardman D.A., Schilling J.W., Miller J., Appleby V.,  
 RA Chen G.C., Kirsner S.W., McIntoe G., Kane J.P.;  
 RT "Isolation of a cDNA clone encoding the amino-terminal region of  
 RT human apolipoprotein B.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 83:1467-1471(1986).  
 [11]  
 RP SEQUENCE OF 1-1670 FROM N.A.  
 RX MEDLINE=86287319; PubMed=3461454;  
 RA Procter A.A., Hardman D.A., Sato K.Y., Schilling J.W.,  
 RA Yamanka M., Hori Y.J., Hjertild K.A., Chen G.C., Kane J.P.;  
 RT "Analysis of cDNA clones encoding the entire B-26 region of human  
 RT apolipoprotein B.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 83:5678-5682(1986).  
 [12]  
 RP PARTIAL SEQUENCE, AND IDENTIFICATION OF APO-B48.  
 RX MEDLINE=88018019; PubMed=3659919;  
 RA Chen S.-H., Habib G., Yang C.-H., Gu Z.-W., Lee B.R., Weng S.-H.,  
 RA Silberman S.R., Cai S.-J., Deslypere J.P., Rosseneu M.,  
 RA Goto A.M., Jr., Li W.-H., Chan L.;  
 RT "Apolipoprotein B-48 is the product of a messenger RNA with an organ-  
 RT specific in-frame stop codon.";  
 RL Science 238:363-366(1987).  
 [13]  
 RP DOMAINS.  
 RX MEDLINE=87039351; PubMed=3773997;  
 RA Knott T.C., Pease R.J., Powell L.M., Wallis S.C., Rall S.C. Jr.,  
 RA Innerarity T.L., Blackhart B., Taylor W.R., Marcel Y., Milne R.,  
 RA Johnson D., Fuller M., Lusis A.J., McCarthy B.J., Mahley R.W.,  
 RA Levy-Wilson B., Scott J.;  
 RT "Complete protein sequence and identification of structural domains  
 RT of human apolipoprotein B.";  
 RL Nature 323:734-738(1986).  
 [14]  
 RP DOMAINS.  
 RA Yang C.-Y., Chen S.-H., Gianturco S.H., Bradley W.A., Sparrow J.T.,  
 RA Tanimura M., Li W.-H., Sparrow D.A., Deloof H., Rosseneu M.,  
 RA Lee F.-S., Gu Z.-W., Goto A.M., Jr., Chan L.;  
 RT "Sequence, structure, receptor-binding domains and internal repeats  
 RT of human apolipoprotein B-100.";  
 RL Nature 323:738-742(1986).  
 [15]  
 RP CALCIUM-BINDING DATA.  
 RX MEDLINE=86242245; PubMed=3087360;  
 RA Daehli N., Lee D.M., Mok T.;  
 RT "Apolipoprotein B is a calcium binding protein.";  
 RL Biochem. Biophys. Res. Commun. 137:493-499(1986).  
 [16]  
 RP PALMITOYLATION OF CY5-1112.  
 RX MEDLINE=20143590; PubMed=10679026;  
 RA Zhao Y., McCabe J.B., Vance J., Berthiaume L.G.;  
 RT "Palmitoylation of apolipoprotein B is required for proper  
 RT intracellular sorting and transport of cholesterol esters and  
 RT triglycerides.";  
 RL Mol. Biol. Cell 11:721-734(2000).  
 [17]  
 RP VARIANT SER-4338.  
 RX MEDLINE=91071750; PubMed=1979313;  
 RA Navajas M., Laurent A.-M., Moreel J.-F., Ragab A., Cambou J.-P.,  
 RA Cuny G., Cambien F., Roizes G.;  
 RT "Detection by denaturing gradient gel electrophoresis of a new  
 RT polymorphism in the apolipoprotein B gene.";  
 RL Hum. Genet. 86:91-93(1990).  
 [18]  
 RP VARIANT FDB GLN-3527.  
 RX MEDLINE=8908975; PubMed=2563166;  
 RA Soria L.F., Ludwig E.H., Clarke H.R.G., Vega G.L., Grundy S.M.,  
 RA McCarthy B.D.;  
 RT "Association between a specific apolipoprotein B mutation and  
 RT familial defective apolipoprotein B-100.";

RL Proc. Natl. Acad. Sci. U.S.A. 86:587-591(1989).  
 RN [19]  
 RP VARIANT LEU-2739.  
 RX MEDLINE=91016974; PubMed=2216805;  
 RA Huang L.-S., Gaviish D., Breslow J.L.;  
 RT "Sequence polymorphism in the human APOB gene at position 8344.";  
 RL Nucleic Acids Res. 18:5922-5922(1990).  
 RN [20]  
 RP VARIANT FDB CYS-3558.  
 RX MEDLINE=95190020; PubMed=7883971;  
 RA Pullinger C.R., Hennessey L.K., Chatterton J.E., Liu W., Love J.A.,  
 RA Mendel C.M., Frost P.H., Malloy M.J., Schumaker V.N., Kane J.P.;  
 RT "Familial ligand-defective apolipoprotein B. Identification of a new  
 RT mutation that decreases LDL receptor binding affinity.";  
 RL J. Clin. Invest. 95:1225-1234(1995).  
 RN [21]  
 RP VARIANT LEU-1437; SER-1914; LYS-2566; THR-3121; ALA-3945; MET-4128  
 RP AND THR-4481.  
 RX MEDLINE=97044521; PubMed=8889592;  
 RA Poirier O., Ricard S., Behague I., Souriau C., Evans A.E.,  
 RA Arveiler D., Marques-Vidal P., Luc G., Roizes G., Cambien F.;  
 RT "Detection of new variants in the apolipoprotein B (apo B) gene by  
 RT PCR-SSCP.";  
 RL Hum. Mutat. 8:282-285(1996).  
 RN [22]  
 RP VARIANT FDB GLN-3527 AND CYS-3558.  
 RX MEDLINE=97403938; PubMed=9259199;  
 RA Rabes J.P., Varret M., Saint-Jore B., Erlich D., Jondeau G.,  
 RA Krempf M., Giraudet P., Junien C., Boileau C.;  
 RT "Familial ligand-defective apolipoprotein B-100: simultaneous  
 RT detection of the ARG3500-->GLN and ARG3531-->CYS mutations in a  
 RT French population.";  
 RL Hum. Mutat. 10:160-163(1997).  
 RN [23]  
 RP VARIANT SER-1914; ARG-1923; LEU-2739; ASP-3319; THR-3427; GLN-3432  
 RP AND ILE-3921.  
 RX MEDLINE=98141125; PubMed=9490296;  
 RA Lereen T.P., Bakken K.S., Hoel V., Hjermann I., Berg K.;  
 RT "Screening for mutations of the apolipoprotein B gene causing  
 RT hypocholesterolemia.";  
 RL Hum. Genet. 102:44-49(1998).  
 CC -1- FUNCTION: APOLIPOPROTEIN B IS A MAJOR PROTEIN CONSTITUENT OF  
 CC CHYLOMICRONS, VLDL AND LDL. IT FUNCTIONS AS A RECOGNITION SIGNAL  
 CC FOR THE CELLULAR BINDING AND INTERNALIZATION OF LDL PARTICLES BY  
 CC THE APOB/E RECEPTOR.

QY Query Match 0.4%; Score 8; DB 1; Length 4563;  
 Best Local Similarity 100.0%; Pred. No. 1e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 1312 GTYGLSCQ 1319  
 DB 1499 GTYGLSCQ 1506

RESULT 10  
 ID RL24\_BACHD STANDARD; PRT; 103 AA.  
 AC Q929K3; Q929K6;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE 50S ribosomal protein L24.  
 GN RPLX OR BH0145.  
 OS Bacillus halodurans.  
 CC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 ON NCBI\_TaxID=86665;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C-125 / JCM 9153;  
 RX MEDLINE=99209008; PubMed=10192928;  
 RA Takami H., Takaki Y., Nakasone K., Hirama C., Inoue A., Horikoshi K.;  
 RT "Sequence analysis of a 32-kb region including the major ribosomal

RT protein gene clusters from alkaliphilic Bacillus sp. strain C-125.";  
 RL Biosci. Biotechnol. Biochem. 63:452-455(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C-125 / JCM 9153;  
 RX MEDLINE=20512582; PubMed=11058132;  
 RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Maui N.,  
 RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,  
 RA Horikoshi K.;  
 RT "Complete genome sequence of the alkaliphilic bacterium Bacillus  
 RT halodurans and genomic sequence comparison with Bacillus subtilis.";  
 RL Nucleic Acids Res. 28:4317-4331(2000).  
 CC -1- FUNCTION: THIS PROTEIN IS FOUND IN THE RIBONUCLEOPROTEIN CORE AND  
 CC IS INVOLVED IN THE EARLY ASSEMBLY OF THE 50S SUBUNIT. IT IS NOT  
 CC INVOLVED IN THE FUNCTIONS OF THE MATURE 50S SUBUNIT  
 CC (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE L24P FAMILY OF RIBOSOMAL PROTEINS.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AB017508; BAA75282.1; -;  
 CC EMBL: AP001507; BAB03864.1; -;  
 CC PIR: T44394; T44394.  
 CC InterPro: IPR005824; KOW.  
 CC InterPro: IPR006646; KOW\_sub.  
 CC InterPro: IPR003256; Ribosomal\_L24.  
 CC InterPro: IPR005825; Ribosomal\_L24\_26.  
 CC Pfam: PF00467; KOW: 1.  
 CC ProDom: PD001677; Ribosomal\_L24; 1.  
 CC SMART: SM00739; KOW: 1.  
 CC TIGRFAMs: TIGR01079; rplX\_bact; 1.  
 CC PROSITE: PS01108; RIBOSOMAL\_L24; 1.  
 CC KMW Ribosomal protein; Complete proteome.  
 CC SEQUENCE 103 AA; 11217 MW; EPCBE26E787A4558 CRC64;

QY Query Match 0.4%; Score 7; DB 1; Length 103;  
 Best Local Similarity 100.0%; Pred. No. 46;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 327 SGDKXK 333  
 DB 13 SGDKXK 19

RESULT 11  
 ID RL24\_BACST STANDARD; PRT; 103 AA.  
 AC P04455;  
 DT 13-AUG-1987 (Rel. 05, Created)  
 DT 13-AUG-1987 (Rel. 05, Last sequence update)  
 DT 01-AUG-1992 (Rel. 23, Last annotation update)  
 DE 50S ribosomal protein L24.  
 GN RPLX.  
 OS Bacillus stearothermophilus.  
 CC Bacteria; Firmicutes; Bacillales; Bacillaceae; Geobacillus.  
 ON NCBI\_TaxID=1422;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=85257681; PubMed=4018095;  
 RA Kimura M., Kimura J., Ashman K.;  
 RT "The complete primary structure of ribosomal proteins L1, L14, L15,  
 RT L23, L24 and L29 from Bacillus stearothermophilus.";  
 RL Eur. J. Biochem. 150:491-497(1985).  
 CC -1- FUNCTION: THIS PROTEIN IS FOUND IN THE RIBONUCLEOPROTEIN CORE AND  
 CC IS INVOLVED IN THE EARLY ASSEMBLY OF THE 50S SUBUNIT. IT IS NOT  
 CC INVOLVED IN THE FUNCTIONS OF THE MATURE 50S SUBUNIT.  
 CC -1- SIMILARITY: BELONGS TO THE L24P FAMILY OF RIBOSOMAL PROTEINS.

```

DR PIR; A02819; RBS24.
DR InterPro; IPR005824; KOW.
DR InterPro; IPR006646; KOW sub.
DR InterPro; IPR003256; Ribosomal_L24.
DR InterPro; IPR005825; Ribosomal_L24_26.
DR Pfam; PF00467; KOW; 1.
DR ProDom; PD001677; Ribosomal_L24; 1.
DR SMART; SM00739; KOW; 1.
DR TIGRFAMs; TIGR01079; rplX_bact; 1.
DR PROSITE; PS01108; RIBOSOMAL_L24; 1.
KW Ribosomal protein.
SQ SEQUENCE 103 AA; 11218 MW; 803E9E3E338A2068 CRC64;

Query Match          0.4%; Score 7; DB 1; Length 103;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      327 SGKDCKG 333
Db      13 SGKDCKG 19

RESULT 12
RL24_BACSU          STANDARD; PRT; 103 AA.
ID      RL24_BACSU
AC      P12876;
DT      01-OCT-1989 (Rel. 12, Created)
DT      01-OCT-1989 (Rel. 12, Last sequence update)
DT      28-FEB-2003 (Rel. 41, Last annotation update)
DE      50S ribosomal protein L24 (BL23) (12 kDa DNA-binding protein) (HPB12).
GN      RPLX.
OS      Bacillus subtilis.
OC      Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX      NCBI_TaxID=1423;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=168;
RX      MEDLINE=90016806; PubMed=2508062;
RA      Henkin T.M., Moon S.H., Matheakis L.C., Nomura M.;
RT      "Cloning and analysis of the spc ribosomal protein operon of Bacillus
RL      subtilis: comparison with the spc operon of Escherichia coli.";
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=168 / JH642, and W23;
RX      MEDLINE=92211327; PubMed=1556555;
RA      Sharp P.W., Nolan N.C., Nl Cholmain N., Devine K.M.;
RT      "DNA sequence variability at the rplX locus of Bacillus subtilis.";
RN      [3]
RP      SEQUENCE FROM N.A.
RC      STRAIN=168 / Marburg;
RX      MEDLINE=96186897; PubMed=8635744;
RA      Suh J.W., Boylan S.A., Oh S.H., Price C.W.;
RT      "Genetic and transcriptional organization of the Bacillus subtilis
RL      spc-alpha region.";
RN      [4]
RP      SEQUENCE FROM N.A.
RC      STRAIN=168;
RX      MEDLINE=98044033; PubMed=9384377;
RA      Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA      Azevedo V., Berrero M.G., Bessieres P., Bolotin A., Borchert S.,
RA      Borries R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA      Broiliet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
RA      Choi S.K., Codani J.J., Connerion I.F., Cummings N.J., Daniel R.A.,
RA      Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerxon P.T.,
RA      Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA      Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA      Ghim S.Y., Glaeser P., Goffeau A., Golightly E.J., Grandi G.,
RA      Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA      Hilbert H., Holstappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA      Joris B., Karamata D., Kasehara Y., Klaerr-Blanchard M., Klein C.,

```

```

RA Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA Kurita K., Lapius A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Paro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Prescan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekwaka A., Seror S.J., Serron P., Shin B.S., Soldo B.,
RA Sorokin A., Taccini E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Yamakoshi A., Tanaka T., Tepstra P., Tognoni A.,
RA Toseo V., Uchiyama S., Vandendol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzengger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasunoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zimetstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis.";
RL Nature 390:249-256(1997).
RN [5]
RP SEQUENCE OF 1-29.
RC STRAIN=168;
RX MEDLINE=94110240; PubMed=82827710;
RA Arnold-Schulz-gahmen B., Salti-Montesanto V., Nguyen J.,
RA Hirschbein L., le Hegarat F.;
RT "The Bacillus subtilis nucleoid-associated protein HPB12 strongly
RT compact DNA.";
RL J. Bacteriol. 176:50-60(1994).
CC -1- FUNCTION: THIS PROTEIN IS FOUND IN THE RIBONUCLEOPROTEIN CORE AND
CC IS INVOLVED IN THE EARLY ASSEMBLY OF THE 50S SUBUNIT. IT IS NOT
CC INVOLVED IN THE FUNCTIONS OF THE MATURE 50S SUBUNIT.
CC -1- SIMILARITY: BELONGS TO THE L24 FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.1sb-sib.ch/announce/
CC or send an email to license@sb-sib.ch).
CC -----
DR EMBL; X15664; CA33702.1; -
DR EMBL; M81748; AAB59023.1; -
DR EMBL; M81749; AAB59024.1; -
DR EMBL; L47971; AAB06810.1; -
DR EMBL; Z99104; CAB11903.1; -
DR PIR; S05993; RBS2B.
DR Subtilist; BG10759; rplX.
DR InterPro; IPR005824; KOW.
DR InterPro; IPR006646; KOW sub.
DR InterPro; IPR003256; Ribosomal_L24.
DR InterPro; IPR005825; Ribosomal_L24_26.
DR Pfam; PF00467; KOW; 1.
DR ProDom; PD001677; Ribosomal_L24; 1.
DR SMART; SM00739; KOW; 1.
DR TIGRFAMs; TIGR01079; rplX_bact; 1.
DR PROSITE; PS01108; RIBOSOMAL_L24; 1.
KW Ribosomal protein; Complete proteome.
FT VARIANT 12 12 I -> T (IN STRAIN W23).
FT CONFLICT 20 20 Q -> E (IN REF. 5).
SQ SEQUENCE 103 AA; 11142 MW; BB635A9C8BD6738F1 CRC64;

Query Match          0.4%; Score 7; DB 1; Length 103;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      327 SGKDCKG 333
Db      13 SGKDCKG 19

RESULT 13
RL24_ECOLI

```

ID RL24\_ECOLI STANDARD; PRT; 103 AA.  
AC P02425; P37438;  
DT 21-JUN-1986 (Rel. 01, Created)  
DT 21-JUN-1986 (Rel. 01, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE 50S ribosomal protein L24.  
GN RPLX OR B3309 OR Z4679 OR ECS4174 OR STM3429.  
OS Escherichia coli.  
OS Escherichia coli O157:H7, and  
OS Salmonella typhimurium.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=562, 83334, 602;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPECIES=E.coli; STRAIN=K12;  
RX MEDLINE=83220807; PubMed=6222285;  
RA Ceretti D.P., Dean D., Davis G.R., Bedwell D.M., Nomura M.;  
RT "The spc ribosomal protein operon of Escherichia coli: sequence and  
RT cotranscription of the ribosomal protein genes and a protein export  
RT gene."  
RL Nucleic Acids Res. 11:2599-2616(1983).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC SPECIES=E.coli; STRAIN=K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blather F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Valdes J., Glasner J.D., Rode C.K., Mayhew G.R.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12."  
RL Science 277:1453-1474(1997).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC SPECIES=E.coli; STRAIN=O157:H7 / EDL933 / ATCC 700927;  
RX Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
RA Poefal G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
RA Grobeck B.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,  
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
RA Welch R.A., Blattner F.R.;  
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."  
RL Nature 409:529-533(2001).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC SPECIES=E.coli; STRAIN=O157:H7 / RIMD 0509952;  
RX MEDLINE=21156231; PubMed=11258796;  
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,  
RA Kubara S., Shiba T., Hattori M., Shinagawa H.;  
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli  
RT O157:H7 and genomic comparison with a laboratory strain K-12."  
RL DNA Res. 8:11-22(2001).  
RN [5]  
RP SEQUENCE.  
RC SPECIES=E.coli; STRAIN=K12;  
RX MEDLINE=80092112; PubMed=331595;  
RA Wittman-Liebold B.;  
RT "Primary structure of protein L24 from the Escherichia coli  
RT ribosome."  
RL FEBS Lett. 108:75-80(1979).  
RN [6]  
RP SEQUENCE OF 85-103 FROM N.A.  
RC SPECIES=E.coli;  
RX MEDLINE=8119903; PubMed=6262737;  
RA Oline P.O., Nomura M.;  
RT "Translational regulation by ribosomal protein S8 in Escherichia  
RT coli: structural homology between RNA binding site and feedback  
RT target on mRNA."  
RL Nucleic Acids Res. 9:1757-1764(1981).  
RN [7]

RP SEQUENCE FROM N.A.  
RC SPECIES=S.typhimurium; STRAIN=L72 / SCSC1412 / ATCC 700720;  
RX MEDLINE=21534948; PubMed=11677609;  
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,  
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,  
RA Ryan B., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,  
RA Waterston R., Wilson R.K.;  
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium  
RT L72."  
RL Nature 413:852-856(2001).  
RN [8]  
RP SEQUENCE OF 86-103 FROM N.A.  
RC SPECIES=S.typhimurium;  
RX MEDLINE=89125589; PubMed=2464692;  
RA Ceretti D.P., Matchakis L.C., Kearney K.R., Yu L., Nomura M.;  
RT "Translational regulation of the spc operon in Escherichia coli.  
RT Identification and structural analysis of the target site for S8  
RT repressor protein."  
RL J. Mol. Biol. 204:309-329(1988).  
RN [9]  
RP MASS SPECTROMETRY.  
RC SPECIES=E.coli; STRAIN=K12 / ATCC 25404;  
RX MEDLINE=99196679; PubMed=10094780;  
RA Arnold R.J., Reilly J.P.;  
RT "Observation of Escherichia coli ribosomal proteins and their  
RT posttranslational modifications by mass spectrometry."  
RL Anal. Biochem. 269:105-112(1999).  
CC -1- FUNCTION: THIS PROTEIN IS FOUND IN THE RIBONUCLEOPROTEIN CORE AND  
CC IS INVOLVED IN THE EARLY ASSEMBLY OF THE 50S SUBUNIT. IT IS NOT  
CC INVOLVED IN THE FUNCTIONS OF THE MATURE 50S SUBUNIT.  
CC -1- MASS SPECTROMETRY: MW=11186.5; METHOD=MALDI.  
CC -1- SIMILARITY: BELONGS TO THE L24P FAMILY OF RIBOSOMAL PROTEINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X01563; CA25716.1; -  
DR EMBL; U18997; AAA58106.1; -  
DR EMBL; AE000408; AAC76334.1; -  
DR EMBL; AE000556; AAG58430.1; -  
DR EMBL; AF002564; BA337587.1; -  
DR EMBL; M10195; AAA24050.1; -  
DR EMBL; AE008857; AAL22292.1; -  
DR EMBL; M36266; AAA27227.1; -  
DR PIR; B85996; B85996.  
DR PIR; F91150; F91150.  
DR PIR; H65123; R5EC24.  
DR BCGene; EG10884; P1X.  
DR StyGene; SG10366; P1X.  
DR InterPro; IPR005824; KOW.  
DR InterPro; IPR006646; KOW sub.  
DR InterPro; IPR003256; Ribosomal\_L24.  
DR InterPro; IPR005825; Ribosomal\_L24\_L24\_26.  
DR Pfam; PF00467; KOW; 1.  
DR ProDom; PD001677; Ribosomal\_L24; 1.  
DR SMART; SM00739; KOW; 1.  
DR TIGRfams; TIGR01079; P1X bact; 1.  
DR PROSITE; PS01108; RIBOSOMAL\_L24; 1.  
DR Ribosomal protein; Complete proteome.  
FT INIT MET 0  
SQ SEQUENCE 103 AA; 11185 MW; 9F4CB2EFF66FAE95 CRC64;  
Query Match 0.4%; Score 7; DB 1; Length 103;  
Best local Similarity 100.0%; Pred. No. 46;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 15 GKDKGR 21

RESULT 14  
RL24\_HAEIN STANDARD; PRT; 103 AA.

AC P44362;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE 50S ribosomal protein L24.  
GN RPLX OR RPL24 OR H10789.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
OC Pasteurellaceae; Haemophilus.  
NCBI\_TaxID=727;

OK NCBI\_TaxID=727;  
RN  
RP SEQUENCE FROM N.A.  
RC STRAIN=Rd / KM20 / ATCC 51907;  
RX MEDLINE=95350630; PubMed=7542800;  
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
RA Scott J.D., Shiley R., Liu L.-I., Glodek A., Kelley J.M.,  
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
RA Uterback T.R., Hanna M.C., Nguyen D.T., Sauder D.M., Brandon R.C.,  
RA Grehm C.L., Friedman J.L., Fuhrmann J.L., Geoghagen N.S.M.,  
RA Venter J.C.;  
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae  
Rd.";  
RL Science 269:496-512(1995).

CC -I- INVOLVED IN THE EARLY ASSEMBLY OF THE 50S SUBUNIT. IT IS NOT  
CC INVOLVED IN THE FUNCTIONS OF THE MATURE 50S SUBUNIT  
CC (BY SIMILARITY).  
CC -I- SIMILARITY: BELONGS TO THE L24P FAMILY OF RIBOSOMAL PROTEINS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U32762; AAC22447.1; -.  
DR TIGR; G64093; G64093.  
DR InterPro; IPR005824; KOW.  
DR InterPro; IPR006646; KOW\_sub.  
DR InterPro; IPR003256; Ribosomal\_L24.  
DR InterPro; IPR005825; Ribosomal\_L24\_26.  
DR Pfam; PF00467; KOW; 1.  
DR ProDom; PD001677; Ribosomal\_L24; 1.  
DR SMART; SM00739; KOW; 1.  
DR TIGRFAMs; TIGR01079; rplX\_bact; 1.  
DR PROSITE; PS01108; RIBOSOMAL\_L24; 1.  
DR Ribosomal protein; Complete proteome.  
SQ SEQUENCE 103 AA; 11285 MW; 3AD490377702F19 CRC64;

Query Match 0.4%; Score 7; DB 1; Length 103;  
Best Local Similarity 100.0%; Pred. No. 46;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 328 GKDKGR 334  
DB 16 GKDKGR 22

RESULT 15  
RL24\_BUCAK

ID RL24\_BUCAK STANDARD; PRT; 104 AA.

AC P46177;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE 50S ribosomal protein L24.  
GN RPLX.  
OS Buchnera aphidicola (subsp. Acyrthosiphon kondoi) (Acyrthosiphon  
OS kondoi symbiotic bacterium).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Buchnera.  
NCBI\_TaxID=42474;  
OK NCBI\_TaxID=42474;

RN  
RP SEQUENCE FROM N.A.  
RC STRAIN=Kuraishi;  
RX MEDLINE=96051390; PubMed=7584036;  
RA Abe R., Yamashita A., Isono K.;  
RT "Cloning and characterization of the ribosomal protein genes in the  
RT spc operon of a prokaryotic endosymbiont of the pea aphid,  
RT Acyrthosiphon kondoi.";  
RL DNA Res. 1:103-114(1994).  
CC -I- INVOLVED IN THE EARLY ASSEMBLY OF THE 50S SUBUNIT. IT IS NOT  
CC INVOLVED IN THE FUNCTIONS OF THE MATURE 50S SUBUNIT  
CC (BY SIMILARITY).  
CC -I- SIMILARITY: BELONGS TO THE L24P FAMILY OF RIBOSOMAL PROTEINS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; D31786; BAA0586.1; -.  
DR InterPro; IPR005824; KOW.  
DR InterPro; IPR006646; KOW\_sub.  
DR InterPro; IPR003256; Ribosomal\_L24.  
DR InterPro; IPR005825; Ribosomal\_L24\_26.  
DR Pfam; PF00467; KOW; 1.  
DR ProDom; PD001677; Ribosomal\_L24; 1.  
DR SMART; SM00739; KOW; 1.  
DR TIGRFAMs; TIGR01079; rplX\_bact; 1.  
DR PROSITE; PS01108; RIBOSOMAL\_L24; 1.  
DR Ribosomal protein.  
SQ SEQUENCE 104 AA; 11336 MW; F3553FE6418BF47C CRC64;

Query Match 0.4%; Score 7; DB 1; Length 104;  
Best Local Similarity 100.0%; Pred. No. 46;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 328 GKDKGR 334  
DB 16 GKDKGR 22

Search completed: January 2, 2004, 16:13:41  
Job time : 23 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: January 2, 2004, 16:10:28 ; Search time 53 Seconds  
(without alignments)

8720.230 Million cell updates/sec

Title: US-09-983-025A-2

Perfect score: 1791

Sequence: 1 MMCLKIRISLAIAGWALC.....AADCDLDECTCRDPKAEHQ 1791

Scoring table: OLIGO

Searched: 830525 seqs, 258052604 residues

Word size : 0

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

SPTREMBL\_23:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phase:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query	Length	ID	Description
1	1791	100.0	1791	4	Q9BXP8 homo sapien
2	1487	83.0	1790	4	Q96PR8 homo sapien
3	1268	70.8	1624	4	Q9H4C9 Q9H4C9
4	708	39.5	826	4	Q96PH7 homo sapien
5	396	22.1	396	4	Q9NUP4 Q9NUP4
6	219	12.2	219	4	Q75997 Q75997
7	40	2.2	1214	11	Q8BUC6 Q8BUC6
8	18	1.0	212	6	Q9SLA1 Q9SLA1
9	18	1.0	213	6	Q9SLA2 Q9SLA2
10	18	1.0	246	6	Q9SLA4 Q9SLA4
11	18	1.0	283	6	Q9SLA3 Q9SLA3
12	11	0.6	1367	11	Q9ES06 Q9ES06
13	11	0.6	1545	11	Q8R4K8 Q8R4K8
14	11	0.6	1574	11	Q8R4K7 Q8R4K7
15	9	0.5	354	11	Q8K423 Q8K423
16	9	0.5	502	16	Q9CGH2 Q9CGH2

17	8	0.4	176	10	Q8LW8 oryza sativ
18	8	0.4	223	3	Q9ZAG5 Q9ZAG5
19	8	0.4	237	2	Q8X000 Q8X000
20	8	0.4	228	10	Q9LW1 arabidopsi
21	8	0.4	223	16	Q8DGH5 synechococ
22	8	0.4	235	10	Q8VY12 arabidopsi
23	8	0.4	252	16	Q8YV63 anabaena sp
24	8	0.4	270	5	Q9VQ75 drosophila
25	8	0.4	275	10	Q8H768 oryza sativ
26	8	0.4	349	16	Q9ZWH3 rhizobium m
27	8	0.4	400	2	Q9L514 psychrobact
28	8	0.4	407	16	Q83904 treponema p
29	8	0.4	457	16	Q97PF6 streptococ
30	8	0.4	457	16	Q8DNV6 streptococ
31	8	0.4	467	5	Q16315 caenorhabd
32	8	0.4	467	5	Q46180 drosophila
33	8	0.4	468	11	Q99LH4 mus musculu
34	8	0.4	468	11	Q8BIR7 mus musculu
35	8	0.4	523	10	Q94BR4 arabidopsi
36	8	0.4	580	10	P93809 arabidopsi
37	8	0.4	620	16	Q8YSY0 anabaena ap
38	8	0.4	675	5	Q46185 drosophila
39	8	0.4	678	5	Q815S0 plasmodium
40	8	0.4	712	16	Q8ZG74 yersinia pe
41	8	0.4	1338	5	Q9GQ8 giardia lam
42	8	0.4	2472	16	Q91696 pseudomonas
43	8	0.4	2477	2	Q87001 pseudomonas
44	8	0.4	3252	4	Q13788 homo sapien
45	8	0.4	4042	2	Q8KT64 photorhabdu

## ALIGNMENTS

### RESULT 1

Q9BXP8	PRELIMINARY;	PRT; 1791 AA.
ID	Q9BXP8	
AC	Q9BXP8;	
DT	01-JUN-2001 (TREMBLrel. 17, Created)	
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)	
DT	01-MAR-2003 (TREMBLrel. 23, Last annotation update)	
DE	Pregnancy-associated plasma preproprotein-A2.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.	
OX	NCBI_TaxID=9606;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	TISSUE=Term placenta;	
RX	MEDLINE=21293000; PubMed=11264294;	
RA	Ovregaard M.T., Boldt H.B., Laursen L.S., Sottirup-Jensen L.,	
RA	Conover C.A., Oxvig C.;	
RT	"Pregnancy-associated Plasma Protein-A2 (PAPP-A2), a Novel Insulin-	
RT	like Growth Factor-binding Protein-5 Proteinase.";	
RL	J. Biol. Chem. 276:21849-21853(2001).	
DR	EMBL: AF311940; AAK31073.1; --	
DR	MEROPS; M46_002;	
DR	Genew; HGNC:14615; PLAC3.	
DR	InterPro; IPR003961; FN_III.	
DR	InterPro; IPR002052; N6_Mtease.	
DR	InterPro; IPR000800; Notch.	
DR	InterPro; IPR000436; Sushi_SCR_CCP.	
DR	InterPro; IPR006025; Zn_MTPeptide.	
DR	Pfam; PF00084; sushi; 4.	
DR	SMART; SM00032; CCP; 4.	
DR	SMART; SM00060; FN3; 1.	
DR	SMART; SM00004; NL; 2.	
DR	PROSITE; PS00092; N6_MTASE; 1.	
DR	PROSITE; PS00142; ZINC_PROTEASE; 1.	
SO	SEQUENCE 1791 AA; 198537 MW; P436030821ECGEDD CRC64;	

Query Match 100.0%; Score 1791; DB 4; Length 1791;  
Best Local Similarity 100.0%; Pred. No. 0;

Matches 1791; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 1 MMCKIIRISLAIAGMALCSANSELGWRKKSIVREHILNOVLEGERCMLGKVRPR 60
Db 1 MMCKIIRISLAIAGMALCSANSELGWRKKSIVREHILNOVLEGERCMLGKVRPR 60
Qy 61 ASPOHLLFQVPSBAGNYLRYVGEORIHHTGSKPDTGNAVASLVPDLTENPAGLRG 120
Db 61 ASPOHLLFQVPSBAGNYLRYVGEORIHHTGSKPDTGNAVASLVPDLTENPAGLRG 120
Qy 121 AVEBPAFVGDSPRIGSELLGDDAYIGNORSKESLCEBAGI QKGSAMAATTTAIFTL 180
Db 121 AVEBPAFVGDSPRIGSELLGDDAYIGNORSKESLCEBAGI QKGSAMAATTTAIFTL 180
Qy 181 NEBPETORRGMAKSRORQYWKRAEDQGDGSGISHFOPMPKSLGHRVKKSPRESN 240
Db 181 NEBPETORRGMAKSRORQYWKRAEDQGDGSGISHFOPMPKSLGHRVKKSPRESN 240
Qy 241 QNGEGSYREAFETNSQVGLPILYFSGRERLLRPEVLAIPREAFTEAMVXREGON 300
Db 241 QNGEGSYREAFETNSQVGLPILYFSGRERLLRPEVLAIPREAFTEAMVXREGON 300
Qy 301 NPATIAGVFDCSHTSVDKGMALGIRSGDKGKRDARFPFSLCTDRVKKATILISHSRQ 360
Db 301 NPATIAGVFDCSHTSVDKGMALGIRSGDKGKRDARFPFSLCTDRVKKATILISHSRQ 360
Qy 361 PGTTHVAATYDGHMAYVDGTQVASSLDOSGPLNSPFMASCSILLGDSSEBQXHF 420
Db 361 PGTTHVAATYDGHMAYVDGTQVASSLDOSGPLNSPFMASCSILLGDSSEBQXHF 420
Qy 421 GHLGLTVWSTALPQSHFQSHSSGEEATDVLATLAFPEVNTWVPFDEKXPRLV 480
Db 421 GHLGLTVWSTALPQSHFQSHSSGEEATDVLATLAFPEVNTWVPFDEKXPRLV 480
Qy 481 LOGEPEBEIISPLQPLCGQTVCDNVELISQYNSWYPLRGEKVIROYVNI CDDEGINP 540
Db 481 LOGEPEBEIISPLQPLCGQTVCDNVELISQYNSWYPLRGEKVIROYVNI CDDEGINP 540
Qy 541 IVSEQRLOHLEALNEAFSRNISMQLSVHOVNSTLHRVVLVNCESKIGNDHDEC 600
Db 541 IVSEQRLOHLEALNEAFSRNISMQLSVHOVNSTLHRVVLVNCESKIGNDHDEC 600
Qy 601 EHPPLGYDGDQRLQGRCYSMNRDGLCHEVCNNMLNFDGDCDDPOADVARKTCFDPD 660
Db 601 EHPPLGYDGDQRLQGRCYSMNRDGLCHEVCNNMLNFDGDCDDPOADVARKTCFDPD 660
Qy 661 SPKRAVSVKELKEALQJNSTHPLNTYFASVREDLAGAATPMDKDAVTHLGIVLSPA 720
Db 661 SPKRAVSVKELKEALQJNSTHPLNTYFASVREDLAGAATPMDKDAVTHLGIVLSPA 720
Qy 721 YYGMRGHTDNIHEGHVGLYHVPKGVSEBSCNDPCKETVPSMETGDLCAADTAPEPKS 780
Db 721 YYGMRGHTDNIHEGHVGLYHVPKGVSEBSCNDPCKETVPSMETGDLCAADTAPEPKS 780
Qy 781 ELCEBEPSTDCGTFPPGAPFTNYSYTDNCTDNTPNOVARMHCYLDLVYQOMTES 840
Db 781 ELCEBEPSTDCGTFPPGAPFTNYSYTDNCTDNTPNOVARMHCYLDLVYQOMTES 840
Qy 841 RKPPPIPIPMVIGQTNKSLTIHMLPISGVYDASGSLCGACTEDGTQFYVHTASSR 900
Db 841 RKPPPIPIPMVIGQTNKSLTIHMLPISGVYDASGSLCGACTEDGTQFYVHTASSR 900
Qy 901 RVCSGSGVTPBEAVGPVDVQPCPSLQAWSPEVHALVHMNTVPCPEGGSLELFFHP 960
Db 901 RVCSGSGVTPBEAVGPVDVQPCPSLQAWSPEVHALVHMNTVPCPEGGSLELFFHP 960
Qy 961 VQADTLTLMVTSFMESSQVLFDETEILENKESVHAGLDTCFDIPLTIKLVHDKVSGV 1020
Db 961 VQADTLTLMVTSFMESSQVLFDETEILENKESVHAGLDTCFDIPLTIKLVHDKVSGV 1020
Qy 1021 KVYTFDERIEIDALLISQPSPLCSGCRPVRYOVLADPPASGLPVVVTSHRKTVE 1080
Db 1021 KVYTFDERIEIDALLISQPSPLCSGCRPVRYOVLADPPASGLPVVVTSHRKTVE 1080

```

```

Qy 1081 VTPGOMYQYOVLAEAGGELGASPPNLNHIGAPYCGDKYSERLAGECDGDLVSGDGS 1140
Db 1081 VTPGOMYQYOVLAEAGGELGASPPNLNHIGAPYCGDKYSERLAGECDGDLVSGDGS 1140
Qy 1141 KVCELEBGNFCVGBPSLCYMBGDGICEPFEKRTSIVDCGIYTPKGYLDQMATRAYSHE 1200
Db 1141 KVCELEBGNFCVGBPSLCYMBGDGICEPFEKRTSIVDCGIYTPKGYLDQMATRAYSHE 1200
Qy 1201 DKKKCPVSLVTGEPHSLICTSYHPDLNHRPLTCMFCVASENETODRSEQEGSLKE 1260
Db 1201 DKKKCPVSLVTGEPHSLICTSYHPDLNHRPLTCMFCVASENETODRSEQEGSLKE 1260
Qy 1261 DEWVLKCFNRPGEARAFITFLTDGLVPEBHOQPTVLTLTVRGSNHSIGTYGLSCQH 1320
Db 1261 DEWVLKCFNRPGEARAFITFLTDGLVPEBHOQPTVLTLTVRGSNHSIGTYGLSCQH 1320
Qy 1321 NPILIINTVTHQNVLFHHTTSVLNLFSSPRVGISAVALRTSSRIGLSAPNSCISEDEGQH 1380
Db 1321 NPILIINTVTHQNVLFHHTTSVLNLFSSPRVGISAVALRTSSRIGLSAPNSCISEDEGQH 1380
Qy 1381 QGOSCHIRPCGKQDSCSLLBDHADVYVNCISIGBGLMKCAITQORFALQASSGQYIRPM 1440
Db 1381 QGOSCHIRPCGKQDSCSLLBDHADVYVNCISIGBGLMKCAITQORFALQASSGQYIRPM 1440
Qy 1441 QKEIILTCSSGMDQNVSCLPVDCGVPDPLVNYANFSCSEGTKFLKRCISICVPPAKLQ 1500
Db 1441 QKEIILTCSSGMDQNVSCLPVDCGVPDPLVNYANFSCSEGTKFLKRCISICVPPAKLQ 1500
Qy 1501 GLSPWLTCLEBDGLMSLEBYVCKLECDAPRIILNANLLPHCLQDNHDVGTICKYECKPGY 1560
Db 1501 GLSPWLTCLEBDGLMSLEBYVCKLECDAPRIILNANLLPHCLQDNHDVGTICKYECKPGY 1560
Qy 1561 YVASESAGKRNLLKIQCEBGGIWEQSGCIPVCEPPEPVFGMECTNGFSLDSQCVL 1620
Db 1561 YVASESAGKRNLLKIQCEBGGIWEQSGCIPVCEPPEPVFGMECTNGFSLDSQCVL 1620
Qy 1621 NCNORERKLPILCTKEGLMTQEFKLCENLQGECPPESELSNYSYKCEQGYIGAVCSPL 1680
Db 1621 NCNORERKLPILCTKEGLMTQEFKLCENLQGECPPESELSNYSYKCEQGYIGAVCSPL 1680
Qy 1681 CVIPSPDPMVLPENITADTLHEHMEPVKQSVICTGRQHPVPLVHCTIQSECEPFGADG 1740
Db 1681 CVIPSPDPMVLPENITADTLHEHMEPVKQSVICTGRQHPVPLVHCTIQSECEPFGADG 1740
Qy 1741 WCDTINNRAYCHYDGGCCSSTLSSKVIIPFADCDLDECTCRDPKAEHQ 1791
Db 1741 WCDTINNRAYCHYDGGCCSSTLSSKVIIPFADCDLDECTCRDPKAEHQ 1791

```

## RESULT 2

```

ID Q96PH8 PRELIMINARY; PRT; 1790 AA.
AC Q96PH8;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Pregnancy-associated plasma protein El.
GN PAPPE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=2148266; PubMed=11597188;
RA Page N.M., Butlin D.J., Lomthaisong K., Lowry P.J.;
RT "The Characterization of Pregnancy Associated Plasma Protein-E and the
RT Identification of an Alternative Splice Variant.";
RL Placenta 22:681-687(2001).
DR EMBL; AF342989; AAL17779.1; -.
DR InterPro; IPR002052; N6 Mcase.
DR InterPro; IPR000800; Notch.

```





NCBI\_TaxID=9606;  
[1]  
RN SEQUENCE FROM N.A.  
RC TISSUE=Placenta;  
RA MEDLINE=20472054; PubMed=11018262;  
FA Fair M., Strube J., Geppert H.G., Kocourek A., Mahne M., Teschaeche H.;  
RT "Pregnancy-associated plasma protein-E (PAP-E).";  
RL Biochim. Biophys. Acta 1493:356-362(2000).  
DR EMBL: AJ278348; CAC1134.1; -.  
DR MEROPS: M46.002; -.  
DR InterPro: IPR003961; FN III.  
DR InterPro: IPR002052; N6\_Mtase.  
DR InterPro: IPR000800; Notch.  
DR InterPro: IPR000436; Sushi\_SCR\_CCP.  
DR InterPro: IPR006025; Zn\_MTpeptidase.  
DR Pfam: PF00084; Sushi; 4.  
DR SMART: SM00032; CCP; 4.  
DR SMART: SM00060; FN3; 1.  
DR SMART: SM00004; NL; 2.  
DR PROSITE: PS00092; N6\_MTASE; 1.  
DR PROSITE: PS00142; ZINC\_PROTEASE; 1.  
SQ SEQUENCE 1624 AA; 160426 MW; 106E2F1F9C3B2CB5 CRC64;  
  
Query Match 70.8%; Score 1268; DB 4; Length 1624;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1568; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
QY 168 MAATTTTATFTTLPNPKRETQRGAKSRQRQVWKKRAEDQGDSDGSISSHFPWPKSL 227  
DB 1 MAATTTTATFTTLPNPKRETQRGAKSRQRQVWKKRAEDQGDSDGSISSHFPWPKSL 60  
QY 228 KHRVKSPPESNONGSGSYREAEFTFNSOYGLPTLYSGRRERLLPEVLAETPREAF 287  
DB 61 KHRVKSPPESNONGSGSYREAEFTFNSOYGLPTLYSGRRERLLPEVLAETPREAF 120  
QY 288 TVEAWVKEGGQNNBAIIAGVFNCSHTVSDKGMALGIRSGDKKRDAREFFSLCTDRV 347  
DB 121 TVEAWVKEGGQNNBAIIAGVFNCSHTVSDKGMALGIRSGDKKRDAREFFSLCTDRV 180  
QY 348 KKATLLISHSRYPQGTWTHVAATYDGRMALVYDGTQVASSISDQGPLNSPFMASCRSL 407  
DB 181 KKATLLISHSRYPQGTWTHVAATYDGRMALVYDGTQVASSISDQGPLNSPFMASCRSL 240  
QY 408 LGGDSSEGGHYFRGLGLTVFWSTALPOSHFOSSQSSGSEAEATDLVLTASFEBVNTW 467  
DB 241 LGGDSSEGGHYFRGLGLTVFWSTALPOSHFOSSQSSGSEAEATDLVLTASFEBVNTW 300  
QY 468 VFPRDEKYPRLLEVLOGFEPPEIILSLPQLPCLQGVCDNVELISQYNGWPLRGEKVI 527  
DB 301 VFPRDEKYPRLLEVLOGFEPPEIILSLPQLPCLQGVCDNVELISQYNGWPLRGEKVI 360  
QY 528 QVWVNCDEEGNLPYISEEQIRLOHEALNEAFSRNYISWQLSVHQVHNSLTRRNVLVNCE 587  
DB 361 QVWVNCDEEGNLPYISEEQIRLOHEALNEAFSRNYISWQLSVHQVHNSLTRRNVLVNCE 420  
QY 588 PSKIGNDHCDEPCHEPLTGYDGDGRLQGRCSNMRDGLCHVENNNLNPDDDCDP 647  
DB 421 PSKIGNDHCDEPCHEPLTGYDGDGRLQGRCSNMRDGLCHVENNNLNPDDDCDP 480  
QY 648 QVADVARTCFPDDSPKRAYMSYKELKEALQLNSTHFLNIYPASSVREDLAQAATPMDXD 707  
DB 481 QVADVARTCFPDDSPKRAYMSYKELKEALQLNSTHFLNIYPASSVREDLAQAATPMDXD 540  
QY 708 AVTHLGGIVLSPAYYGMGHDTMTHEVGHVGLYHVKGVSESSCNDPCKETVPSMET 767  
DB 541 AVTHLGGIVLSPAYYGMGHDTMTHEVGHVGLYHVKGVSESSCNDPCKETVPSMET 600  
QY 768 GDLGADTAPTPKSELCRBEPTSDTCGFRPGAFPTMYMSTTDNCTDNFTPNQVAMH 827  
DB 601 GDLGADTAPTPKSELCRBEPTSDTCGFRPGAFPTMYMSTTDNCTDNFTPNQVAMH 660  
QY 828 CYLDLVYQWMTESRKPTPIPPMVIQGTNKSLLTIHMLPPISGVYDRAAGSLGACTED 887

DB 661 CYLDLVYQWMTESRKPTPIPPMVIQGTNKSLLTIHMLPPISGVYDRAAGSLGACTED 720  
QY 888 GTFROYVHTASSRRVDCSSGYMTPEEAVGPDPDQCEBLSQAMSEVHLVHNMVPCP 947  
DB 721 GTFROYVHTASSRRVDCSSGYMTPEEAVGPDPDQCEBLSQAMSEVHLVHNMVPCP 780  
QY 948 TEGCSLELFOHVQADTLTLMWTSFPMSSQVLPTEILLENKESVHLGAPLDPFCDIPL 1007  
DB 781 TEGCSLELFOHVQADTLTLMWTSFPMSSQVLPTEILLENKESVHLGAPLDPFCDIPL 840  
QY 1008 TIKLHVDGKYGKVVYTFDERISIDALLTSQHSPLCSGCRVRRQVLRDPFASGLPV 1067  
DB 841 TIKLHVDGKYGKVVYTFDERISIDALLTSQHSPLCSGCRVRRQVLRDPFASGLPV 900  
QY 1068 VVTHSHRKFTDVTPEQWQYQVLAAGSELBAASPPLNHIGAPYCGDGKYSERLGE 1127  
DB 901 VVTHSHRKFTDVTPEQWQYQVLAAGSELBAASPPLNHIGAPYCGDGKYSERLGE 960  
QY 1128 CDDGDLVSGDGSKVCLEBEGFNCVGEPSLCVMEYEGDLCEPERKTSIVDCGIYTPKGY 1187  
DB 961 CDDGDLVSGDGSKVCLEBEGFNCVGEPSLCVMEYEGDLCEPERKTSIVDCGIYTPKGY 1020  
QY 1188 LDMATRAYSHEDKKKCPVSLVTGEPHSLICTSYHDPLENNRPLTGMFPCVASENETOP 1247  
DB 1021 LDMATRAYSHEDKKKCPVSLVTGEPHSLICTSYHDPLENNRPLTGMFPCVASENETOP 1080  
QY 1248 DRSQPRGSLKKDEWLVKVCFNRPGEARAIPLFTLDGIVPEBHOPTVTLTLTVRGS 1307  
DB 1081 DRSQPRGSLKKDEWLVKVCFNRPGEARAIPLFTLDGIVPEBHOPTVTLTLTVRGS 1140  
QY 1308 NNSLGTGYGLCOHNPFLINTVTHQNVLFHHTSVLNFSSPRVIGSAVALRTSSRIGLSA 1367  
DB 1141 NNSLGTGYGLCOHNPFLINTVTHQNVLFHHTSVLNFSSPRVIGSAVALRTSSRIGLSA 1200  
QY 1368 PSNCISDEQNHQSGCIHRPCGKODSCPSLLIDHADVNTCSIGPGLMKCAITCQRGF 1427  
DB 1201 PSNCISDEQNHQSGCIHRPCGKODSCPSLLIDHADVNTCSIGPGLMKCAITCQRGF 1260  
QY 1428 ALQASSGOYIRPMOKETLLTCSSGHMDQNVSCIPVOCGVDPBLVNVANPSCSEGRKFLK 1487  
DB 1261 ALQASSGOYIRPMOKETLLTCSSGHMDQNVSCIPVOCGVDPBLVNVANPSCSEGRKFLK 1320  
QY 1488 RCSIISCVPAKLQGLSFWLTCLEBGLWSLPEVYCKLECDAPRIILNANLLPQLDQNH 1547  
DB 1321 RCSIISCVPAKLQGLSFWLTCLEBGLWSLPEVYCKLECDAPRIILNANLLPQLDQNH 1380  
QY 1548 VGTTCYKTECKRGIYVAESAAGKVRNKLKIQCLBGGIWEQSGCIPIVCEPPPVFBGMYE 1607  
DB 1381 VGTTCYKTECKRGIYVAESAAGKVRNKLKIQCLBGGIWEQSGCIPIVCEPPPVFBGMYE 1440  
QY 1608 CTNGFSLDSCVLCNOCERERKPLTCTKBLMTQEPFLCENLOGECPSPSELSVYKC 1667  
DB 1441 CTNGFSLDSCVLCNOCERERKPLTCTKBLMTQEPFLCENLOGECPSPSELSVYKC 1500  
QY 1668 EOGYIGAVCSPLCVIIPSDPVMLENITADTLEHNMVYKVQVSIYCTGRQWHPDVLV 1727  
DB 1501 EOGYIGAVCSPLCVIIPSDPVMLENITADTLEHNMVYKVQVSIYCTGRQWHPDVLV 1560  
QY 1728 HGIQSCPEFOA 1738  
DB 1561 HGIQSCPEFOA 1571  
  
RESULT 4  
Q96PH7 PRELIMINARY; PRT; 826 AA.  
AC Q96PH7;  
DT 01-DEC-2001 (Tremblere1.19, Created)  
DT 01-DEC-2001 (Tremblere1.19, Last sequence update)  
DT 01-MAR-2003 (Tremblere1.23, Last annotation update)  
DE Pregnancy-associated plasma protein E2.  
GN PAPPE.  
OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=2148266; PubMed=11597188;  
 RA Page N.M., Butlin D.J., Lomthaisong K., Lowry P.J.;  
 RT "The Characterization of Pregnancy Associated Plasma Protein-E and the  
 RT Identification of an Alternative Splice Variant."  
 RL Placenta 22:681-687(2001)  
 DR EMBL; AF342990; AAL17780.1; -  
 DR InterPro; IPR000800; Notch.  
 DR InterPro; IPR006025; Zn\_MTPetase.  
 DR SMART; SM00004; NL; 1.  
 DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 SQ SEQUENCE 826 AA; 92032 MW; FBF3A09D552065D1 CRC64;

Query Match 39.5%; Score 708; DB 4; Length 826;  
 Best Local Similarity 99.9%; Pred. No. 0;  
 Matches 808; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 MCLTILRISLAILAGMALCSANSELGWRKKSIVREHLNOVLEGERCWTGATVRPRA 61  
 DB 1 MCLTILRISLAILAGMALCSANSELGWRKKSIVREHLNOVLEGERCWTGATVRPRA 60  
 QY 62 SPQHLFGVPSRAGNYLRPYVGEQEIHTGRSKPTDEGNASVLPEDLTENPAGLRGA 121  
 DB 61 SPQHLFGVPSRAGNYLRPYVGEQEIHTGRSKPTDEGNASVLPEDLTENPAGLRGA 120  
 QY 122 VEEPAAPWGDSPIGQSEELGDDDAVYLGNGRSKESLGAGIQKGSAMAAITTTAIFTTLN 181  
 DB 121 VEEPAAPWGDSPIGQSEELGDDDAVYLGNGRSKESLGAGIQKGSAMAAITTTAIFTTLN 180  
 QY 182 EPRKETQRGAKRQRRQVWKRAEDQGSGISHSNQMPKPSLKRRVKSPPESNQ 241  
 DB 181 EPRKETQRGAKRQRRQVWKRAEDQGSGISHSNQMPKPSLKRRVKSPPESNQ 240  
 QY 242 NGSGSRREATFNSQVGLPILYFSGRERLLRPREVAEIPREAFTEAVWKEGGONN 301  
 DB 241 NGSGSRREATFNSQVGLPILYFSGRERLLRPREVAEIPREAFTEAVWKEGGONN 300  
 QY 302 PAIAGVFNCSHTVSDKGALGIRSGDKGRDARPPFSLCTDRVKKATILISHRYOP 361  
 DB 301 PAIAGVFNCSHTVSDKGALGIRSGDKGRDARPPFSLCTDRVKKATILISHRYOP 360  
 QY 362 GTWTHVAATYGRHMAIYVDGTQVASSLDQSGPLNSPFMASCRSLILGDSSEDEGHYFRG 421  
 DB 361 GTWTHVAATYGRHMAIYVDGTQVASSLDQSGPLNSPFMASCRSLILGDSSEDEGHYFRG 420  
 QY 422 HLGTLVFWSTALPQSHFQSSQSSGEEATDVLTLASFEFVNTWVWPFREKYPRLLEV 481  
 DB 421 HLGTLVFWSTALPQSHFQSSQSSGEEATDVLTLASFEFVNTWVWPFREKYPRLLEV 480  
 QY 482 QGFEPPELISPLQPLCGQTVCDNVELISQYNGWPLRGEKVIKYOVNVCDEGLNPI 541  
 DB 481 QGFEPPELISPLQPLCGQTVCDNVELISQYNGWPLRGEKVIKYOVNVCDEGLNPI 540  
 QY 542 VSEEQIRLOHEALNEAFSRNYSWQLSVHQNSTLRHRVVLVNCPEKIGNDHCEPE 601  
 DB 541 VSEEQIRLOHEALNEAFSRNYSWQLSVHQNSTLRHRVVLVNCPEKIGNDHCEPE 600  
 QY 602 HPLTGYGGDGRCLQRCYSNMRDGLCHVECNMNLNDPDDGCDPOVADVRKTCFDPDS 661  
 DB 601 HPLTGYGGDGRCLQRCYSNMRDGLCHVECNMNLNDPDDGCDPOVADVRKTCFDPDS 660  
 QY 662 PKRAYMSYKELKEALQLNSTHFLNIYFASVREDLAGATPMDKADATVHLGATVSPAY 721  
 DB 661 PKRAYMSYKELKEALQLNSTHFLNIYFASVREDLAGATPMDKADATVHLGATVSPAY 720  
 QY 722 YGMNGHTDTMHEGVHVTGLYHVFKEGVERBSNDPCKEYTPSNETGDLCDTAPTPKSE 781  
 DB 721 YGMNGHTDTMHEGVHVTGLYHVFKEGVERBSNDPCKEYTPSNETGDLCDTAPTPKSE 780

QY 782 LCREPPTSQGTFRPGAPFTNMSYT 810  
 DB 781 LCREPPTSQGTFRPGAPFTNMSYT 809

RESULT 5  
 Q9NUF4 PRELIMINARY; PRT; 396 AA.  
 ID Q9NUF4  
 AC Q9NUF4  
 DT 01-OCT-2000 (Tremblrel. 15, Created)  
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE D65218.1 (Similar to PAPA (Pregnancy-associated plasma protein A))  
 DE (Fragment).  
 GN D65218.1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Donnelly S.;  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL031734; CAB72297.1; -  
 DR MEROPS; M46.002; -  
 DR InterPro; IPR000800; Notch.  
 DR SMART; SM00004; NL; 1.  
 FT NON\_TER 1  
 SQ SEQUENCE 396 AA; 44619 MW; CD8C1BF3B8097506 CRC64;

Query Match 22.1%; Score 396; DB 4; Length 396;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 396; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 269 RERLLRPEVLAIEPREAFTEAVWKEGGONNPAIAGVFNCSHTVSDKGALGIRSG 328  
 DB 1 RERLLRPEVLAIEPREAFTEAVWKEGGONNPAIAGVFNCSHTVSDKGALGIRSG 60  
 QY 329 KDKGKRARPPFSLCTDRVKKATILISHRYOPETWTHVAATYDGRHMAIYVDGTQVASS 388  
 DB 61 KDKGKRARPPFSLCTDRVKKATILISHRYOPETWTHVAATYDGRHMAIYVDGTQVASS 120  
 QY 389 LDQSGPLNSPFMASCRSLILGDSSEDEGHYFRGLGTLVFWSTALPQSHFQSSQSSGE 448  
 DB 121 LDQSGPLNSPFMASCRSLILGDSSEDEGHYFRGLGTLVFWSTALPQSHFQSSQSSGE 180  
 QY 449 BEATDVLTLASFEFVNTWVWPFREKYPRLLEVQGFEPPELISPLQPLCGQTVCDNVE 508  
 DB 181 BEATDVLTLASFEFVNTWVWPFREKYPRLLEVQGFEPPELISPLQPLCGQTVCDNVE 240  
 QY 509 LISQYNGWPLRGEKVIKYOVNVCDEGLNPIVSEEQIRLOHEALNEAFSRNYSWQLS 568  
 DB 241 LISQYNGWPLRGEKVIKYOVNVCDEGLNPIVSEEQIRLOHEALNEAFSRNYSWQLS 300  
 QY 569 VHOVHNSTLRHRVVLVNCPEKIGNDHCEPEHPLTGYGGDGRCLQRCYSNMRDGLC 628  
 DB 301 VHOVHNSTLRHRVVLVNCPEKIGNDHCEPEHPLTGYGGDGRCLQRCYSNMRDGLC 360  
 QY 629 HVECNMNLNDPDDGCDPOVADVRKTCFDPDSBKR 664  
 DB 361 HVECNMNLNDPDDGCDPOVADVRKTCFDPDSBKR 396

RESULT 6  
 O75997 PRELIMINARY; PRT; 219 AA.  
 ID O75997  
 AC O75997;  
 DT 01-NOV-1998 (Tremblrel. 08, Created)  
 DT 01-JAN-1999 (Tremblrel. 09, Last sequence update)  
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
 DE D0774124.1 (Similar to pregnancy-associated plasma protein A precursor) (Fragment).  
 GN D0774124.1.

OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Grifham D.;  
 RL Submitted (OCT-1998) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AL031290; CAZ0358.1; -.  
 DR InterPro; IPR000800; Notch.  
 DR InterPro; IPR000436; Sushi\_SCR\_CCP.  
 DR SMART; SM00032; CCP; 1.  
 DR SMART; SM00004; NL; 1.  
 FT NON TER 1  
 SQ SEQUENCE 219 AA; 24366 MW; 65C5311409C759DA CRC64;

Query Match 12.2%; Score 219; DB 4; Length 219;  
 Best Local Similarity 100.0%; Pred. No. 4.4e-224;  
 Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1573 KLLKIQEGGIWEGSCIPVCEPPVFEQVYECTNGFSLDSCVINCNOERKLPIL 1632  
 DB 1 KLLKIQEGGIWEGSCIPVCEPPVFEQVYECTNGFSLDSCVINCNOERKLPIL 60  
 OY 1633 CTKEGLMTQEFKLCENLQGECPPEPSELNSVEYKCEQYIGAVCSPLCVIPSPDVLMP 1692  
 DB 61 CTKEGLMTQEFKLCENLQGECPPEPSELNSVEYKCEQYIGAVCSPLCVIPSPDVLMP 120  
 OY 1693 ENITADTLEHMMWEPKVSIVCTGRQWHPDPVLVHCIQSCFPADGMCOTINRAYCH 1752  
 DB 121 ENITADTLEHMMWEPKVSIVCTGRQWHPDPVLVHCIQSCFPADGMCOTINRAYCH 180  
 OY 1753 YDGGDCSSTLSKKVIFPAADCDUDECCTCRDPKAEENQ 1791  
 DB 181 YDGGDCSSTLSKKVIFPAADCDUDECCTCRDPKAEENQ 219

## RESULT 7

OY 08BUG6 PRELIMINARY; PRT; 1214 AA.  
 ID 08BUG6  
 AC 08BUG6  
 DT 01-MAR-2003 (TREMBlrel. 23, Created)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE Similar to pregnancy-associated plasma PREPROTEIN-A2 (Fragment).  
 OS Mus musculus (Mouse).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Scltognathi; Muridae; Murinae; Mus.  
 NX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=C57BL/6J; TISSUE=Dorsal root ganglion;  
 RX MEDLINE=22354683; PubMed=12466851;  
 RA The FANTOM Consortium;  
 RT the RIKEN Genome Exploration Research Group Phase I & II Team;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs."  
 RL Nature 420:563-573(2002).  
 DR EMBL; AK084006; BAC39093.1; -.  
 FT NON TER 1  
 FT NON TER 1  
 SQ SEQUENCE 1214 AA; 134983 MW; AF4204160F05A285 CRC64;

Query Match 2.2%; Score 40; DB 11; Length 1214;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-32;  
 Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 608 DGGDCRLQGRCYSNRBDGLCHVECNMMLNDPDDCCDP 647  
 DB 446 DGGDCRLQGRCYSNRBDGLCHVECNMMLNDPDDCCDP 485

## RESULT 8

OY 095L41 PRELIMINARY; PRT; 212 AA.  
 ID 095L41  
 AC 095L41  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE Pregnancy-associated plasma protein-A (Fragment).  
 GN PAPP-A.  
 OS Equus caballus (Horse).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.  
 NX NCBI\_TaxID=9796;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Mazerbourg S.; Overgaard M.T.; Oxvig C.; Christiansen M.;  
 RA Conover C.A.; Laurendau I.; Vidaud M.; Tosser-Klopp G.; Zapf J.;  
 RA Monget P.;  
 RT "Pregnancy Associated Plasma Protein-A (PAPP-A) in ovine, bovine,  
 RT porcine and equine ovarian follicles: involvement in IGFBP-4  
 RT proteolytic degradation and mRNA expression during follicular  
 RT development."  
 RL Endocrinology 0:0-0(2001).  
 DR EMBL; AF421143; AAL16087.1; -.  
 FT NON TER 1  
 FT NON TER 1  
 SQ SEQUENCE 212 AA; 23516 MW; 351475A848E0EAB CRC64;

Query Match 1.0%; Score 18; DB 6; Length 212;  
 Best Local Similarity 100.0%; Pred. No. 7.7e-10;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 818 FTPNOVARMHGCLDLYVO 835  
 DB 61 FTPNOVARMHGCLDLYVO 78

## RESULT 9

OY 095L42 PRELIMINARY; PRT; 213 AA.  
 ID 095L42  
 AC 095L42  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE Pregnancy-associated plasma protein-A (Fragment).  
 GN PAPP-A.  
 OS Sus scrofa (Pig).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.  
 NX NCBI\_TaxID=9823;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Mazerbourg S.; Overgaard M.T.; Oxvig C.; Christiansen M.;  
 RA Conover C.A.; Laurendau I.; Vidaud M.; Tosser-Klopp G.; Zapf J.;  
 RA Monget P.;  
 RT "Pregnancy Associated Plasma Protein-A (PAPP-A) in ovine, bovine,  
 RT porcine and equine ovarian follicles: involvement in IGFBP-4  
 RT proteolytic degradation and mRNA expression during follicular  
 RT development."  
 RL Endocrinology 0:0-0(2001).  
 DR EMBL; AF421142; AAL16086.1; -.  
 FT NON TER 1  
 FT NON TER 1  
 SQ SEQUENCE 213 AA; 23684 MW; D1205E8445324A8C CRC64;

Query Match 1.0%; Score 18; DB 6; Length 213;  
 Best Local Similarity 100.0%; Pred. No. 7.7e-10;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 818 FTPNOVARMHGCLDLYVO 835  
 DB 61 FTPNOVARMHGCLDLYVO 78

```

RESULT 10
Q95L44 PRELIMINARY; PRT; 246 AA.
ID Q95L44;
AC Q95L44;
DT 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE Pregnancy-associated plasma protein-A (Fragment).
GN PAPP-A.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RA Maerbourg S., Overgaard M.T., Oxvig C., Christensen M.,
RA Conover C.A., Laurendau I., Vidau M., Tosser-Klopp G., Zapf J.,
RA Monget P.;
RT "Pregnancy Associated Plasma Protein-A (PAPP-A) in ovine, bovine,
RT porcine and equine ovarian follicles: involvement in IGFBP-4
RT proteolytic degradation and mRNA expression during follicular
RT development.";
RL Endocrinology 0:0-0(2001).
DR EMBL; AF421140; AAL16084.1; -.
FT NON_TER 1 1
FT NON_TER 246 246
SQ SEQUENCE 246 AA; 27275 MW; 1BFC11D34C4E9F07 CRC64;

Query Match 1.0%; Score 18; DB 6; Length 246;
Best Local Similarity 100.0%; Pred. No. 8,7e-10;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 818 FTPNOVARMHCYLDLVYQ 835
Db 86 FTPNOVARMHCYLDLVYQ 103

RESULT 11
Q95L43 PRELIMINARY; PRT; 283 AA.
ID Q95L43;
AC Q95L43;
DT 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE Pregnancy-associated plasma protein-A (Fragment).
GN PAPP-A.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA Maerbourg S., Overgaard M.T., Oxvig C., Christensen M.,
RA Conover C.A., Laurendau I., Vidau M., Tosser-Klopp G., Zapf J.,
RA Monget P.;
RT "Pregnancy Associated Plasma Protein-A (PAPP-A) in ovine, bovine,
RT porcine and equine ovarian follicles: involvement in IGFBP-4
RT proteolytic degradation and mRNA expression during follicular
RT development.";
RL Endocrinology 0:0-0(2001).
DR EMBL; AF421141; AAL16085.1; -.
FT NON_TER 1 1
FT NON_TER 283 283
SQ SEQUENCE 283 AA; 31243 MW; EEE461A616F39333 CRC64;

Query Match 1.0%; Score 18; DB 6; Length 283;
Best Local Similarity 100.0%; Pred. No. 9,9e-10;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 818 FTPNOVARMHCYLDLVYQ 835

```

```

Db 86 FTPNOVARMHCYLDLVYQ 103

RESULT 12
Q9ES06 PRELIMINARY; PRT; 1367 AA.
ID Q9ES06;
AC Q9ES06;
DT 01-MAR-2001 (TREMblrel. 16, Created)
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
DE Pregnancy-associated plasma protein-A (Fragment).
GN PAPP-A.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Hourvitz A., Kuwahara A., Hennebold J.D., Negishi H., Tanaka M.,
RA Widge A., Erickson G.P., Adashi E.Y.;
RT "The regulated expression of insulin-like growth factor binding
RT protein-4-endorpeptidase (Pregnancy-associated plasma protein-A) in the
RT rodent ovary: A marker of the dominant follicle and of the corpus
RT luteum.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF258461; AAC09799.1; -.
DR MEROPS; M46.001; -.
DR MGD; MGI:97479; Pappa.
DR InterPro; IPR000800; Notch.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR InterPro; IPR006025; Zn_MTpeptidase.
DR Pfam; PF00084; Sushi; 5.
DR SMART; SM00032; CCP; 4.
DR SMART; SM00004; NL; 3.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
FT NON_TER 1 1
FT NON_TER 1367 1367
SQ SEQUENCE 1367 AA; 152644 MW; A46BC5B0449042A5 CRC64;

Query Match 0.6%; Score 11; DB 11; Length 1367;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 825 RMHCYLDLVYQ 835
Db 394 RMHCYLDLVYQ 404

RESULT 13
Q8R4K8 PRELIMINARY; PRT; 1545 AA.
ID Q8R4K8;
AC Q8R4K8;
DT 01-JUN-2002 (TREMblrel. 21, Created)
DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)
DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
DE Pregnancy-associated plasma protein-A (Fragment).
GN PAPP-A.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Soe R., Overgaard M.T., Thomsen A.R., Laursen J.S., Olsen I.M.,
RA Haaning J., Sottrup-Jensen L., Haaning J., Giudice L.C., Conover C.A.,
RA Oxvig C.;
RT "Expression of recombinant murine PAPP-A and a novel variant (PAPP-A1)
RT with differential proteolytic activity.";
RL Eur. J. Biochem. 0:0-0(2002).
DR EMBL; AF439513; AAM12687.1; -.
DR MGD; MGI:97479; Pappa.
DR InterPro; IPR006558; LamG_like.

```

DR InterPro; IPR000800; Notch.  
 DR InterPro; IPR000436; Sush1\_SCR\_CCP.  
 DR InterPro; IPR006025; Zn\_Mtpeptidase.  
 DR Pfam; PF00084; sush1; 5.  
 DR SMART; SM00032; CCP; 4.  
 DR SMART; SM00560; LamG; 1.  
 DR SMART; SM00004; NL; 3.  
 DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 DR NON\_TER 1  
 SQ SEQUENCE 1545 AA; 172584 MW; B820BF998245419 CRC64;

Query Match 0.6%; Score 11; DB 11; Length 1545;  
 Best Local Similarity 100.0%; Pred. No. 0.12;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 825 RMHCYLDLVYQ 835  
 DB 572 RMHCYLDLVYQ 582

## RESULT 14

Q8R4K7 PRELIMINARY; PRT; 1574 AA.  
 AC Q8R4K7;  
 DT 01-JUN-2002 (TREMBlrel. 21, Created)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE Pregnancy-associated plasma protein-A variant (Fragment).  
 GN PAPPA.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Soe R., Overgaard M.T., Thomsen A.R., Laursen L.S., Olsen I.M.,  
 RA Haaning J., Solttrup-Jensen L., Haaning J., Gludice L.C., Conover C.A.,  
 RA Oxyvig C.;  
 RT "Expression of recombinant murine PAPPA and a novel variant (PAPPA-I)  
 RT with differential proteolytic activity.";  
 RL Eur. J. Biochem. 0:0-0(2002).  
 DR EMBL; AF439514; AAM12688.1; -.  
 DR MGI; MGI:97479; Pappa.  
 DR InterPro; IPR006558; LamG\_like.  
 DR InterPro; IPR000800; Notch.  
 DR InterPro; IPR000436; Sush1\_SCR\_CCP.  
 DR InterPro; IPR006025; Zn\_Mtpeptidase.  
 DR Pfam; PF00084; sush1; 5.  
 DR SMART; SM00032; CCP; 4.  
 DR SMART; SM00560; LamG; 1.  
 DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 DR NON\_TER 1  
 SQ SEQUENCE 1574 AA; 176105 MW; AE53F0187E6EFB7B CRC64;

Query Match 0.6%; Score 11; DB 11; Length 1574;  
 Best Local Similarity 100.0%; Pred. No. 0.12;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 825 RMHCYLDLVYQ 835  
 DB 601 RMHCYLDLVYQ 611

## RESULT 15

Q8K423 PRELIMINARY; PRT; 354 AA.  
 AC Q8K423;  
 DT 01-OCT-2002 (TREMBlrel. 22, Created)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE Pregnancy-associated plasma protein A (Fragment).  
 GN PAPPA.

OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=129/SvJ;  
 RA Qin X., Sexton C., Byun D., Strong D.D., Baylink D.J., Mohan S.;  
 RT "Differential Regulation of Pregnancy Associated Plasma Protein  
 RT (PAPPA) during Pregnancy in Human and Mouse.";  
 RL Growth Horm. IGF Res. 0:0-0(2002).  
 DR EMBL; AF510317; AAM44048.1; -.  
 DR InterPro; IPR006558; LamG\_like.  
 DR InterPro; IPR000800; Notch.  
 DR SMART; SM00560; LamG; 1.  
 DR SMART; SM00004; NL; 2.  
 DR NON\_TER 1  
 FT NON\_TER 1  
 FT 354  
 SQ SEQUENCE 354 AA; 40550 MW; 06093321EF32805 CRC64;

Query Match 0.5%; Score 9; DB 11; Length 354;  
 Best Local Similarity 100.0%; Pred. No. 4.7;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 654 KTCFDPDSP 662  
 DB 345 KTCFDPDSP 353

Search completed: January 2, 2004, 16:14:48  
 Job time : 56 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

# OM protein - protein search, using sw model

Run on: January 2, 2004, 16:12:13 ; Search time 26 Seconds  
(without alignments)  
2914.567 Million cell updates/sec

Title: US-09-983-025a-2

Sequence: 1 MMCKLIRISLAIAGMALC.....AADCDDECTCRDPAEENG 1791

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 328717 seqs, 42310858 residues

Word size : 0

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Issued Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/1aa/5A COMB pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B COMB pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A COMB pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B COMB pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PCUS COMB pep:\*  
6: /cgn2\_6/ptodata/1/1aa/backfile1.ppe.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8	0.4	105	US-09-732-210-781	Sequence 781, App
2	8	0.4	400	US-09-252-991A-31900	Sequence 31900, A
3	8	0.4	422	US-08-190-802A-53	Sequence 53, Appl
4	8	0.4	422	US-08-190-802A-55	Sequence 55, Appl
5	8	0.4	422	US-08-477-346-53	Sequence 53, Appl
6	8	0.4	422	US-08-477-346-55	Sequence 55, Appl
7	8	0.4	422	US-08-473-089-53	Sequence 53, Appl
8	8	0.4	422	US-08-473-089-55	Sequence 55, Appl
9	8	0.4	422	US-08-487-072A-53	Sequence 53, Appl
10	8	0.4	422	US-08-487-072A-55	Sequence 55, Appl
11	8	0.4	592	US-09-536-059-3	Sequence 3, Appl
12	8	0.4	4536	US-09-180-422B-27	Sequence 27, Appl
13	8	0.4	4536	US-09-180-006C-1	Sequence 1, Appl
14	7	0.4	18	US-08-920-610-4	Sequence 4, Appl
15	7	0.4	18	US-09-140-149-2	Sequence 2, Appl
16	7	0.4	18	US-08-672-213-4	Sequence 4, Appl
17	7	0.4	18	US-08-973-131-31	Sequence 31, Appl
18	7	0.4	18	US-09-615-917-2	Sequence 2, Appl
19	7	0.4	92	US-08-942-423-9	Sequence 9, Appl
20	7	0.4	28	US-09-107-532A-4945	Sequence 4945, Ap
21	7	0.4	103	US-09-732-210-762	Sequence 762, App
22	7	0.4	103	US-09-732-210-763	Sequence 763, App
23	7	0.4	103	US-09-732-210-765	Sequence 765, App
24	7	0.4	103	US-09-732-210-766	Sequence 766, App
25	7	0.4	104	US-09-732-210-761	Sequence 761, App
26	7	0.4	107	US-09-370-838-113	Sequence 113, App
27	7	0.4	108	US-09-252-991A-21351	Sequence 21351, A

28	7	0.4	113	US-09-732-210-774	Sequence 774, App
29	7	0.4	115	US-09-732-210-780	Sequence 780, App
30	7	0.4	144	US-09-252-991A-16576	Sequence 16576, A
31	7	0.4	151	US-09-328-352-7958	Sequence 7958, Ap
32	7	0.4	187	US-09-352-991A-28834	Sequence 28834, A
33	7	0.4	188	US-09-134-001C-4249	Sequence 4249, Ap
34	7	0.4	203	US-09-252-991A-17282	Sequence 17282, A
35	7	0.4	209	US-08-827-279-1	Sequence 1, Appl
36	7	0.4	209	US-09-345-473B-6	Sequence 6, Appl
37	7	0.4	219	US-09-107-532A-5199	Sequence 5199, Ap
38	7	0.4	222	US-09-328-352-4436	Sequence 4436, Ap
39	7	0.4	232	US-09-252-991A-29073	Sequence 29073, A
40	7	0.4	241	US-09-328-352-5920	Sequence 5920, Ap
41	7	0.4	244	US-09-252-991A-21937	Sequence 21937, A
42	7	0.4	245	US-09-858-664A-7	Sequence 7, Appl
43	7	0.4	247	US-09-252-991A-32400	Sequence 32400, A
44	7	0.4	286	US-09-252-991A-32062	Sequence 32062, A
45	7	0.4	291	US-09-252-991A-28755	Sequence 28755, A

## ALIGNMENTS

```

RESULT 1
US-09-732-210-781
; Sequence 781, Application US/09732210
; Patent No. 6573361
; GENERAL INFORMATION:
; APPLICANT: Bunkers, Greg J.
; APPLICANT: Liang, Jihong
; APPLICANT: Mitanck, Cindy A.
; APPLICANT: Seale, Jeffrey W.
; APPLICANT: Wu, Yonnie S.
; TITLE OF INVENTION: Anti-Serum Proteins and Methods for Their Use
; FILE REFERENCE: 38-21(15036) B
; CURRENT APPLICATION NUMBER: US/09/732,210
; CURRENT FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/169,513
; PRIOR FILING DATE: 1999-12-07
; PRIOR APPLICATION NUMBER: US 60/169,340
; PRIOR FILING DATE: 1999-12-07
; NUMBER OF SEQ ID NOS: 1753
; SEQ ID NO 781
; LENGTH: 105
; TYPE: PRT
; ORGANISM: Thermotoga maritima
US-09-732-210-781

Query Match      0.4% Score 8; DB 4; Length 105;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      327 SGKDKGR 334
Db      13 SGKDKGR 20

RESULT 2
US-09-252-991A-31900
; Sequence 31900, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142

```

SEQ ID NO 31900  
LENGTH: 400  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-31900

Query Match 0.4%; Score 8; DB 4; Length 400;  
Best Local Similarity 100.0%; Pred. No. 36;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 278 VLAEPR 285  
Db 189 VLAEPR 196

RESULT 3  
US-08-190-802A-53  
Sequence 53, Application US/08190802A  
Patent No. 5519003

GENERAL INFORMATION:  
APPLICANT: Mochly-Rosen, Daria  
TITLE OF INVENTION: MD-40 - Derived Peptides and Uses  
NUMBER OF SEQUENCES: 265  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dehlinger & Associates  
STREET: P.O. Box 60850  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94306-0850

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/190,802A  
FILING DATE: 01-FEB-1994  
CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:  
NAME: Fabian, Gary R.  
REGISTRATION NUMBER: 33,875  
REFERENCE/DOCKET NUMBER: 8600-0139  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 324-0960  
TELEFAX: (415) 324-0960  
INFORMATION FOR SEQ ID NO: 53:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 422 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
INDIVIDUAL ISOLATE: MS11, Fig. 36  
US-08-190-802A-53

Query Match 0.4%; Score 8; DB 1; Length 422;  
Best Local Similarity 100.0%; Pred. No. 38;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 TEGNAVSL 106  
Db 271 TEGNAVSL 278

RESULT 4  
US-08-190-802A-55  
Sequence 55, Application US/08190802A  
Patent No. 5519003

GENERAL INFORMATION:  
APPLICANT: Mochly-Rosen, Daria  
TITLE OF INVENTION: MD-40 - Derived Peptides and Uses  
NUMBER OF SEQUENCES: 265  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dehlinger & Associates  
STREET: P.O. Box 60850  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94306-0850

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/190,802A  
FILING DATE: 01-FEB-1994  
CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:  
NAME: Fabian, Gary R.  
REGISTRATION NUMBER: 33,875  
REFERENCE/DOCKET NUMBER: 8600-0139  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 324-0960  
TELEFAX: (415) 324-0960  
INFORMATION FOR SEQ ID NO: 55:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 422 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
INDIVIDUAL ISOLATE: ORF R81, Fig. 38  
US-08-190-802A-55

Query Match 0.4%; Score 8; DB 1; Length 422;  
Best Local Similarity 100.0%; Pred. No. 38;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 TEGNAVSL 106  
Db 271 TEGNAVSL 278

RESULT 5  
US-08-477-346-53  
Sequence 53, Application US/08477346  
Patent No. 6262023

GENERAL INFORMATION:  
APPLICANT: Mochly-Rosen, Daria  
TITLE OF INVENTION: MD-40 - Derived Peptides and Uses  
NUMBER OF SEQUENCES: 265  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Morison & Foerster  
STREET: 2000 Pennsylvania Avenue, NW  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20006-1812

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:



APPLICATION NUMBER: US/08/477,346  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/487,072  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: MURASHIGE, KATE H.  
REGISTRATION NUMBER: 29,959  
REFERENCE/DOCKET NUMBER: 2550-0025.20  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 887-1500  
TELEFAX: (202) 887-0763  
INFORMATION FOR SEQ ID NO: 53:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 422 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
INDIVIDUAL ISOLATE: MSL1, Fig. 36  
US-08-477-346-53

Query Match 0.4%; Score 8; DB 3; Length 422;  
Best Local Similarity 100.0%; Pred. No. 38;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 TEGNAVSL 106  
Db 271 TEGNAVSL 278

RESULT 6  
US-08-477-346-55  
Sequence 55, Application US/08477346  
Patent No. 6262023  
GENERAL INFORMATION:  
APPLICANT: Mochly-Rosen, Daria  
TITLE OF INVENTION: WD-40 - Derived Peptides and Uses  
TITLE OF INVENTION: Thee of  
NUMBER OF SEQUENCES: 265  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Morrison & Foerster  
STREET: 2000 Pennsylvania Avenue, NW  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20006-1812  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/477,346  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/487,072  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: MURASHIGE, KATE H.  
REGISTRATION NUMBER: 29,959  
REFERENCE/DOCKET NUMBER: 2550-0025.20  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 887-1500  
TELEFAX: (202) 887-0763  
INFORMATION FOR SEQ ID NO: 55:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 422 amino acids

TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
INDIVIDUAL ISOLATE: ORF RB1, Fig. 38  
US-08-477-346-55

Query Match 0.4%; Score 8; DB 3; Length 422;  
Best Local Similarity 100.0%; Pred. No. 38;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 TEGNAVSL 106  
Db 271 TEGNAVSL 278

RESULT 7  
US-08-473-089-53  
Sequence 53, Application US/08473089  
Patent No. 6342368  
GENERAL INFORMATION:  
APPLICANT: Mochly-Rosen, Daria  
APPLICANT: Ron, Dorit  
TITLE OF INVENTION: WD-40 - Derived Peptides and Uses  
TITLE OF INVENTION: Thee of  
NUMBER OF SEQUENCES: 265  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Morrison & Foerster  
STREET: 2000 Pennsylvania Avenue, NW  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20006-1812  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/473,089  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: MURASHIGE, KATE H.  
REGISTRATION NUMBER: 29,959  
REFERENCE/DOCKET NUMBER: 2550-0025.22  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 887-1500  
TELEFAX: (202) 887-0763  
INFORMATION FOR SEQ ID NO: 53:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 422 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
INDIVIDUAL ISOLATE: MSL1, Fig. 36  
US-08-473-089-53

Query Match 0.4%; Score 8; DB 4; Length 422;  
Best Local Similarity 100.0%; Pred. No. 38;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 TEGNAVSL 106  
Db 271 TEGNAVSL 278

RESULT 8

US-08-473-089-55  
Sequence 55, Application US/08473089  
Patent No. 642368  
GENERAL INFORMATION:  
APPLICANT: Mochly-Rosen, Daria  
ATTORNEY/AGENT INFORMATION:  
NAME: MURASHIGE, KATE H.  
REGISTRATION NUMBER: 29,959  
REFERENCE/DOCKET NUMBER: 2550-0025.22  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 887-0763  
TELEFAX: (202) 887-1500  
INFORMATION FOR SEQ ID NO: 55:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 422 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
INDIVIDUAL ISOLATE: ORF RB1, Fig. 38  
US-08-473-089-55

Query Match 0.4%; Score 8; DB 4; Length 422;  
Best Local Similarity 100.0%; Pred. No. 38;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 TEGNAVSL 106  
Db 271 TEGNAVSL 278

RESULT 9  
US-08-487-072A-53  
Sequence 53, Application US/08487072A  
Patent No. 642364  
GENERAL INFORMATION:  
APPLICANT: Mochly-Rosen, Daria  
ATTORNEY/AGENT INFORMATION:  
NAME: MURASHIGE, KATE H.  
REGISTRATION NUMBER: 29,959  
REFERENCE/DOCKET NUMBER: 2550-0025.20  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 887-0763  
TELEFAX: (202) 887-1500  
INFORMATION FOR SEQ ID NO: 53:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 422 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/487,072A  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: MURASHIGE, KATE H.  
REGISTRATION NUMBER: 29,959  
REFERENCE/DOCKET NUMBER: 2550-0025.20  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 887-0763  
TELEFAX: (202) 887-1500  
INFORMATION FOR SEQ ID NO: 53:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 422 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein

US-08-487-072A-53  
Sequence 53, Application US/08487072A  
Patent No. 642364  
GENERAL INFORMATION:  
APPLICANT: Mochly-Rosen, Daria  
ATTORNEY/AGENT INFORMATION:  
NAME: MURASHIGE, KATE H.  
REGISTRATION NUMBER: 29,959  
REFERENCE/DOCKET NUMBER: 2550-0025.20  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 887-0763  
TELEFAX: (202) 887-1500  
INFORMATION FOR SEQ ID NO: 53:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 422 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein

Query Match 0.4%; Score 8; DB 4; Length 422;  
Best Local Similarity 100.0%; Pred. No. 38;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 TEGNAVSL 106  
Db 271 TEGNAVSL 278

RESULT 10  
US-08-487-072A-55  
Sequence 55, Application US/08487072A  
Patent No. 6423684  
GENERAL INFORMATION:  
APPLICANT: Mochly-Rosen, Daria  
ATTORNEY/AGENT INFORMATION:  
NAME: MURASHIGE, KATE H.  
REGISTRATION NUMBER: 29,959  
REFERENCE/DOCKET NUMBER: 2550-0025.20  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 887-0763  
TELEFAX: (202) 887-1500  
INFORMATION FOR SEQ ID NO: 55:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 422 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein

HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
INDIVIDUAL ISOLATE: ORF RB1, Fig. 38  
US-08-487-072A-55

Query Match 0.4%; Score 8; DB 4; Length 422;  
Best Local Similarity 100.0%; Pred. No. 38;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 TEGNAVSL 106  
DB 271 TEGNAVSL 278

RESULT 11  
US-09-536-059-3  
Sequence 3, Application US/09536059  
Patent No. 6544737  
GENERAL INFORMATION:  
APPLICANT: Blumenfeld, Marta  
APPLICANT: Chumakov, Ilya  
APPLICANT: Bougueterec, Lydie  
APPLICANT: Cohen-Avalene, Annick  
TITLE OF INVENTION: GENOMIC SEQUENCE OF THE PURH GENE AND PURH-RELATED BIALLELIC  
FILE REFERENCE: GENSET.058AUS  
CURRENT APPLICATION NUMBER: US/09/536,059  
CURRENT FILING DATE: 2000-03-31  
PRIOR APPLICATION NUMBER: US 60/125,961  
PRIOR FILING DATE: 1999-03-24  
NUMBER OF SEQ ID NOS: 24  
SOFTWARE: Patent.pm  
SEQ ID NO: 3  
LENGTH: 592  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: VARIANT  
LOCATION: 116  
OTHER INFORMATION: Xaa=Thr or Ser  
US-09-536-059-3

Query Match 0.4%; Score 8; DB 4; Length 592;  
Best Local Similarity 100.0%; Pred. No. 52;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 669 VKELKEAL 676  
DB 250 VKELKEAL 257

RESULT 12  
US-09-180-422B-27  
Sequence 27, Application US/09180422B  
Patent No. 6444644  
GENERAL INFORMATION:  
APPLICANT: BRUCKDORFER, KARL R  
ETTELME, CAMILLE  
TITLE OF INVENTION: ANTICOAGULANT PEPTIDE FRAGMENTS DERIVED  
FROM APOLIPROTEIN B-100  
NUMBER OF SEQUENCES: 27  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NIXON & VANDERHAYE, P.C.  
STREET: 1100 NORTH GLEBE ROAD, 8TH FLOOR  
CITY: ARLINGTON  
STATE: VA  
COUNTRY: USA  
ZIP: 22201  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/180,422B  
FILING DATE: 07-Dec-1998  
CLASSIFICATION: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: SADOFF, B. J.  
REGISTRATION NUMBER: 36663  
REFERENCE/DOCKET NUMBER: 117-268  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 7038164000  
TELEFAX: 7038164100  
INFORMATION FOR SEQ ID NO: 27:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4536 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 27:  
US-09-180-422B-27

Query Match 0.4%; Score 8; DB 4; Length 4536;  
Best Local Similarity 100.0%; Pred. No. 3.4e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1312 GTYGLSCQ 1319  
DB 1472 GTYGLSCQ 1479

RESULT 13  
US-09-108-006C-1  
Sequence 1, Application US/09108006C  
Patent No. 6524613  
GENERAL INFORMATION:  
APPLICANT: Steer, Clifford J.  
Kren, Betsy T.  
Bandopadhyay, Paramita  
Roy-Chowdhury, Jayanta  
TITLE OF INVENTION: Hepatocellular Chimeraplasty  
NUMBER OF SEQUENCES: 62  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Kimeragen, Inc.  
STREET: 300 Pleasant Run  
CITY: Newtown  
STATE: PA  
COUNTRY: USA  
ZIP: 18940  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/108,006C  
FILING DATE: 30-Jun-1992  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/054,288  
FILING DATE: 30-APR-1997  
APPLICATION NUMBER: 60/054,837  
FILING DATE: 05-AUG-1997  
APPLICATION NUMBER: 60/064,996  
FILING DATE: 10-NOV-1997  
APPLICATION NUMBER: 60/074,497  
FILING DATE: 12-FEB-1998  
APPLICATION NUMBER: PCT US 98/08834  
FILING DATE: 30-APR-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Friebe, Thomas  
REGISTRATION NUMBER: 29258  
REFERENCE/DOCKET NUMBER: 7991-015-999

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-504-4444  
TELEFAX: 215-504-4545  
TELEX: <Unknown>  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4563 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-108-006C-1

Query Match 0.4%; Score 8; DB 4; Length 4563;  
Best Local Similarity 100.0%; Pred. No. 3.4e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1312 GTYGLSCQ 1319  
Db 1499 GTYGLSCQ 1506

RESULT 14  
US-08-920-610-4  
Sequence 4, Application US/08920610  
Patent No. 6015709  
GENERAL INFORMATION:  
APPLICANT: Natesan, Sridaran  
TITLE OF INVENTION: TRANSCRIPTIONAL ACTIVATORS, AND  
TITLE OF INVENTION: COMPOSITIONS AND USES RELATED THERETO  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FOLEY, HOAG & ELIOT LLP  
STREET: One Post Office Square  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02109-2170  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/920,610  
FILING DATE: 27-AUG-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Vincent, Matthew P.  
REGISTRATION NUMBER: 36,709  
REFERENCE/DOCKET NUMBER: APV-006.02  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-832-1000  
TELEFAX: 617-832-7000  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 18 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-920-610-4

Query Match 0.4%; Score 7; DB 3; Length 18;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1034 ALLTSOP 1040  
Db 12 ALLTSOP 18

RESULT 15  
US-09-140-149-2  
Sequence 2, Application US/09140149  
Patent No. 6117680  
GENERAL INFORMATION:  
APPLICANT: Natesan, Sridaran  
TITLE OF INVENTION: No. 6117680el Compositions and Methods for Regulation of  
TITLE OF INVENTION: Transcription  
FILE REFERENCE: 363C  
CURRENT APPLICATION NUMBER: US/09/140,149  
CURRENT FILING DATE: 1998-08-26  
EARLIER APPLICATION NUMBER: 08/918,401  
EARLIER FILING DATE: 1997-08-26  
EARLIER APPLICATION NUMBER: 08/920,610  
EARLIER FILING DATE: 1997-08-27  
EARLIER APPLICATION NUMBER: 09/126,009  
EARLIER FILING DATE: 1998-07-29  
EARLIER APPLICATION NUMBER: PCT/US97/15219  
EARLIER FILING DATE: 1997-08-27  
NUMBER OF SEQ ID NOS: 22  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 2  
LENGTH: 18  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-140-149-2

Query Match 0.4%; Score 7; DB 3; Length 18;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1034 ALLTSOP 1040  
Db 12 ALLTSOP 18

Search completed: January 2, 2004, 16:16:07  
Job time : 27 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2004, 16:14:54 ; Search time 51 Seconds

(without alignments)  
7046.047 Million cell updates/sec

Title: US-09-983-025A-2

Sequence: 1 MMCKTIRISLAIAGMALC.....AACDDDECTCRDRAENQ 1791

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 733937 seqs, 200641211 residues

Word size : 0

Total number of hits satisfying chosen parameters: 733937

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

1: /cgn2\_6/ptodata/1/pubppa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/1/pubppa/US06\_PUBCOMB.pep:\*  
3: /cgn2\_6/ptodata/1/pubppa/US05\_PUBCOMB.pep:\*  
4: /cgn2\_6/ptodata/1/pubppa/US04\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubppa/US03\_PUBCOMB.pep:\*  
6: /cgn2\_6/ptodata/1/pubppa/US02\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/1/pubppa/US01\_PUBCOMB.pep:\*  
8: /cgn2\_6/ptodata/1/pubppa/US00\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/1/pubppa/US09\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/1/pubppa/US08\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/1/pubppa/US07\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/1/pubppa/US06\_PUBCOMB.pep:\*  
13: /cgn2\_6/ptodata/1/pubppa/US05\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/1/pubppa/US04\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/1/pubppa/US03\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/1/pubppa/US02\_PUBCOMB.pep:\*  
17: /cgn2\_6/ptodata/1/pubppa/US01\_PUBCOMB.pep:\*  
18: /cgn2\_6/ptodata/1/pubppa/US00\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1791	100.0	1791	11	US-09-983-025-2
2	1536	85.8	1791	10	US-09-983-025-2
3	1532	85.5	1770	10	US-09-827-998-10
4	574	32.0	1365	10	US-09-827-998-16
5	192	10.7	192	9	US-09-864-761-34265
6	70	3.9	70	9	US-09-864-761-34264
7	63	3.5	63	9	US-09-864-761-34262
8	19	1.1	19	10	US-09-827-998-7
9	18	1.0	1627	11	US-09-983-025-25
10	18	1.0	1627	12	US-10-295-027-663
11	11	0.6	20	10	US-09-827-998-18
12	9	0.5	502	12	US-10-369-493-18401
13	8	0.4	20	15	US-10-115-072-7
14	8	0.4	28	12	US-10-029-386-30575
15	8	0.4	164	14	US-10-078-929-84

16	8	0.4	176	12	US-10-259-165-86	Sequence 86, Appl
17	8	0.4	176	12	US-10-259-165-346	Sequence 346, Appl
18	8	0.4	176	12	US-10-078-929-86	Sequence 86, Appl
19	8	0.4	280	12	US-10-259-165-174	Sequence 174, Appl
20	8	0.4	422	12	US-10-369-493-1483	Sequence 1483, Ap
21	8	0.4	457	9	US-09-815-242-13478	Sequence 13478, A
22	8	0.4	470	15	US-10-156-761-8854	Sequence 8854, Ap
23	8	0.4	481	15	US-10-156-761-14941	Sequence 14941, A
24	8	0.4	489	15	US-10-156-761-8688	Sequence 8688, Ap
25	8	0.4	592	12	US-10-376-893-3	Sequence 3, Appl1
26	8	0.4	598	9	US-09-925-301-1218	Sequence 1218, Ap
27	8	0.4	598	15	US-10-106-698-4589	Sequence 4589, Ap
28	8	0.4	1076	12	US-10-369-493-13831	Sequence 13831, A
29	8	0.4	1460	15	US-10-128-714-8301	Sequence 8301, Ap
30	8	0.4	1706	9	US-09-864-761-46862	Sequence 46862, A
31	8	0.4	2472	9	US-09-815-242-5064	Sequence 5064, Ap
32	8	0.4	4563	10	US-09-870-759-1128	Sequence 128, Ap
33	8	0.4	4563	11	US-09-802-640-32	Sequence 32, Appl
34	8	0.4	4563	12	US-09-751-708A-128	Sequence 128, Ap
35	7	0.4	12	11	US-09-954-385-338	Sequence 338, Ap
36	7	0.4	18	11	US-09-852-370-31	Sequence 31, Appl
37	7	0.4	18	12	US-10-002-244-4	Sequence 4, Appl1
38	7	0.4	18	15	US-10-097-800B-3	Sequence 3, Appl1
39	7	0.4	20	15	US-10-225-567A-1927	Sequence 5, Appl1
40	7	0.4	23	9	US-09-759-287A-5	Sequence 1927, Ap
41	7	0.4	51	12	US-10-029-386-33913	Sequence 33913, A
42	7	0.4	51	11	US-09-764-891-4736	Sequence 4736, Ap
43	7	0.4	58	12	US-10-029-386-30928	Sequence 30928, A
44	7	0.4	71	12	US-10-371-099-347	Sequence 347, Appl
45	7	0.4	71	12	US-10-371-264-51	Sequence 51, Appl

#### ALIGNMENTS

RESULT 1  
US-09-983-025-2  
; Sequence 2, Application US/09983025  
; Publication No. US20030124529A1  
GENERAL INFORMATION:  
; APPLICANT: OXVIG, Claus  
; TITLE OF INVENTION: PREGNANCY-ASSOCIATED PLASMA PROTEIN-A2 (PAPP-A2)  
; FILE REFERENCE: OXVIG-1A  
; CURRENT APPLICATION NUMBER: US/09/983, 025  
; PRIOR FILING DATE: 2001-10-22  
; PRIOR APPLICATION NUMBER: US 60/241,840  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: DK PA 2000 01571  
; NUMBER OF SEQ ID NOS: 25  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2  
; LENGTH: 1791  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: (1)..(66)  
; OTHER INFORMATION: prepro part of PAPP-A2  
; NAME/KEY: misc.feature  
; LOCATION: (67)..(699)  
; OTHER INFORMATION: pro part of PAPP-A2  
; US-09-983-025-2

Query Match 100.0%; Score 1791; DB 11; Length 1791;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1791; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 MMCKTIRISLAIAGMALCSANSEIGWTRKSLVEREHINQVLLBGERCWLAKVRRPR 60  
DB 1 MMCKTIRISLAIAGMALCSANSEIGWTRKSLVEREHINQVLLBGERCWLAKVRRPR 60

QY 61 ASPOHLLFGVYPSRAGNYLREYPVGEQEIHHHTGRSKPDTEGNAVSLVPPDLTENPAGLRG 120  
DB 61 ASPOHLLFGVYPSRAGNYLREYPVGEQEIHHHTGRSKPDTEGNAVSLVPPDLTENPAGLRG 120  
QY 121 AVEBPAPWVDSTIGQSELLGDDAYIGNORSKSLGEGAGIQGASAMAATTTTALIFTL 180  
DB 121 AVEBPAPWVDSTIGQSELLGDDAYIGNORSKSLGEGAGIQGASAMAATTTTALIFTL 180  
QY 181 NEPEPEORRGMASRORROYKRRAPEDQGDGSGISSHFOPMRHSLSKRRKSPREESN 240  
DB 181 NEPEPEORRGMASRORROYKRRAPEDQGDGSGISSHFOPMRHSLSKRRKSPREESN 240  
QY 241 QNGEGSYREAFETNSQVGLPILYFSGRERLLLRPEVLAIRPREAFTVEAMVPEEGON 300  
DB 241 QNGEGSYREAFETNSQVGLPILYFSGRERLLLRPEVLAIRPREAFTVEAMVPEEGON 300  
QY 301 NPALIAGVFDCNSHTVSDKMGALGIRSGKDKRDAFFSLCTDRVKKATILISHSRQ 360  
DB 301 NPALIAGVFDCNSHTVSDKMGALGIRSGKDKRDAFFSLCTDRVKKATILISHSRQ 360  
QY 361 PGTHVAATYDGRHMLLYDGTQVASSLDQSGPLNSPFMASCRSLIGDSDSDGHYR 420  
DB 361 PGTHVAATYDGRHMLLYDGTQVASSLDQSGPLNSPFMASCRSLIGDSDSDGHYR 420  
QY 421 GHLGTLVFWSTALPQSHFOHSSSGEEBATDLVLTASFEVNTWVPPFDEKYPRLV 480  
DB 421 GHLGTLVFWSTALPQSHFOHSSSGEEBATDLVLTASFEVNTWVPPFDEKYPRLV 480  
QY 481 LOGEPEBELISLOPLCGQTCVNDNELISQNGYMPLRGEKYYRQVNI CDDEGLNP 540  
DB 481 LOGEPEBELISLOPLCGQTCVNDNELISQNGYMPLRGEKYYRQVNI CDDEGLNP 540  
QY 541 IYSEEOJRLQHEALNEAFSRNISMOLSVHVNSTLRHRVVLNCEPSKIGNDHCPBC 600  
DB 541 IYSEEOJRLQHEALNEAFSRNISMOLSVHVNSTLRHRVVLNCEPSKIGNDHCPBC 600  
QY 601 EHPPLTGDGDCRLQGRCYSWNRBDGLCHVECNMMLDFDGDCCDPVADVKTCEPD 660  
DB 601 EHPPLTGDGDCRLQGRCYSWNRBDGLCHVECNMMLDFDGDCCDPVADVKTCEPD 660  
QY 661 SPKAYNSVKEALKALQUNSTHPLNIYFASVREDLGAATWPMDCXAVTLGGLVSPA 720  
DB 661 SPKAYNSVKEALKALQUNSTHPLNIYFASVREDLGAATWPMDCXAVTLGGLVSPA 720  
QY 721 YGMRGHTDMIHVGHVGLYHVFEGYSEESGNDPCKEIVPMEGTDLCAADTAPTPKS 780  
DB 721 YGMRGHTDMIHVGHVGLYHVFEGYSEESGNDPCKEIVPMEGTDLCAADTAPTPKS 780  
QY 781 ELCEPEPTSDTCGTFEPGAPFTNYSYTDNCTDNFTPNQVARMHCYLDLVYQOMTES 840  
DB 781 ELCEPEPTSDTCGTFEPGAPFTNYSYTDNCTDNFTPNQVARMHCYLDLVYQOMTES 840  
QY 841 RKPPPIPIPPVNIQOTNKSLLTIHMLPISGVYVIRASGSLGACTEDGTFRQYVHTASSR 900  
DB 841 RKPPPIPIPPVNIQOTNKSLLTIHMLPISGVYVIRASGSLGACTEDGTFRQYVHTASSR 900  
QY 901 RVCSSGMYPEEAIVGPPDVQPCPSIQAMSPEVHLVHMNTVPCPEEGSLELLPOHP 960  
DB 901 RVCSSGMYPEEAIVGPPDVQPCPSIQAMSPEVHLVHMNTVPCPEEGSLELLPOHP 960  
QY 961 VOADTLTLMWTSFMESSQVLFDEILLENKESVHLGLDFTCDIPLTIKLVGDKVSGV 1020  
DB 961 VOADTLTLMWTSFMESSQVLFDEILLENKESVHLGLDFTCDIPLTIKLVGDKVSGV 1020  
QY 1021 KVTTFDERIEIDALLTSQHSPLCSGCRPVRYVLDPPASGLPVVYTHSHRKTDFVE 1080  
DB 1021 KVTTFDERIEIDALLTSQHSPLCSGCRPVRYVLDPPASGLPVVYTHSHRKTDFVE 1080  
QY 1081 VTPGOMYQOVLARAGELGASPLNHHIGAAYCGGKXSERIGEEDDDDLVSGGCS 1140  
DB 1081 VTPGOMYQOVLARAGELGASPLNHHIGAAYCGGKXSERIGEEDDDDLVSGGCS 1140  
QY 1141 KVCELEEGFNCVGEPSLCYMEGDCGICPEFERKTSIVDCGIYTPKGYLDQWATRAYSSHE 1200

DB 1141 KVCELEEGFNCVGEPSLCYMEGDCGICPEFERKTSIVDCGIYTPKGYLDQWATRAYSSHE 1200  
QY 1201 DKKKCPVSLVTGEPHSLICTSYHDPDLPHNRPLTGMPCVASENETQDDBREOEGSLJKE 1260  
DB 1201 DKKKCPVSLVTGEPHSLICTSYHDPDLPHNRPLTGMPCVASENETQDDBREOEGSLJKE 1260  
QY 1261 DEWVLKCFNRPGBARAIFFLTTDGLVPEHQOPTYTLVLTVRGSHSLGTYSQCH 1320  
DB 1261 DEWVLKCFNRPGBARAIFFLTTDGLVPEHQOPTYTLVLTVRGSHSLGTYSQCH 1320  
QY 1321 NPILINVTQHONYLFHHTTSVLNFSFPRVISAVALRTSSRIGLAPSNCSISEDEQNH 1380  
DB 1321 NPILINVTQHONYLFHHTTSVLNFSFPRVISAVALRTSSRIGLAPSNCSISEDEQNH 1380  
QY 1381 QGGSCHRPCKGKDCSLLDHDADVNCISIGPLMKCAITQORFALQASSGQYTRPM 1440  
DB 1381 QGGSCHRPCKGKDCSLLDHDADVNCISIGPLMKCAITQORFALQASSGQYTRPM 1440  
QY 1441 QKEILTLCSGHDONVSCLPVDCGVPDPSLVNANFSCSEGTKFLKRCISICVPPAKQ 1500  
DB 1441 QKEILTLCSGHDONVSCLPVDCGVPDPSLVNANFSCSEGTKFLKRCISICVPPAKQ 1500  
QY 1501 GLSPWLTLCEBDGLMSLPEVYCKLECDAPRIILNANLLPHCLDNHDVGTICKYCKRGY 1560  
DB 1501 GLSPWLTLCEBDGLMSLPEVYCKLECDAPRIILNANLLPHCLDNHDVGTICKYCKRGY 1560  
QY 1561 YVASEAGKVRNLTCLKQCLEGGIWEQSGCIPVCEPFPVFEQMECTNGFSLDSCVL 1620  
DB 1561 YVASEAGKVRNLTCLKQCLEGGIWEQSGCIPVCEPFPVFEQMECTNGFSLDSCVL 1620  
QY 1621 NQNOERKLPILCTKEGLMTQEFKLCENLQCECPPEPSLNSVYKCEQGYIGAVCSPL 1680  
DB 1621 NQNOERKLPILCTKEGLMTQEFKLCENLQCECPPEPSLNSVYKCEQGYIGAVCSPL 1680  
QY 1681 CVIPSDPVLNENITADTLEHMMEPVKQSIYCTGRQHPDPVLVHCIOGCEPQADG 1740  
DB 1681 CVIPSDPVLNENITADTLEHMMEPVKQSIYCTGRQHPDPVLVHCIOGCEPQADG 1740  
QY 1741 WCDTINRAACHYDGGCCSSTSSKVIIPFADCDLDECTCRDPRAENQ 1791  
DB 1741 WCDTINRAACHYDGGCCSSTSSKVIIPFADCDLDECTCRDPRAENQ 1791

RESULT 2  
US-09-827-998-3  
Sequence 3, Application US/09827998  
Patent No. US2002010252A1  
GENERAL INFORMATION:  
APPLICANT: Shannon, Mark  
TITLE OF INVENTION: NOVEL ISOFORMS OF HUMAN PREGNANCY-ASSOCIATED PROTEIN E  
FILE REFERENCE: MDHMR-8  
CURRENT APPLICATION NUMBER: US/09/827,998  
PRIOR FILING DATE: 2001-04-06  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
NUMBER OF SEQ ID NOS: 1881  
SOFTWARE: Aeonica Sequence Listing Engine  
SEQ ID NO 3  
LENGTH: 1791  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-827-998-3

Query Match 85.84; Score 1536; DB 10; Length 1791;  
Best Local Similarity 99.94; Pred. No. 0;  
Matches 1736; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
1 MMCKILIRISLAIAGALCSANSELGWRKSLVEREHLNOVLDEGERCMGAKYRRPR 60

```

Db      1 MMCKILIRISLAILAGALCSANSELGWTTRKSLVEREHINOVILBGERCMLGAKVRRPR 60
Qy      61 ASPOHILFGVYPSRAGNYLRPYVGEBOIHHTGRSKPDTEGNAYSLVPPDLTENPAGLRG 120
Db      61 ASPOHILFGVYPSRAGNYLRPYVGEBOIHHTGRSKPDTEGNAYSLVPPDLTENPAGLRG 120
Qy      121 AVEBPAAWVGDSPITQSGELLGDDDAVLGNQSRKESLGEAGIOKGSMAATTTTATFTTL 180
Db      121 AVEBPAAWVGDSPITQSGELLGDDDAVLGNQSRKESLGEAGIOKGSMAATTTTATFTTL 180
Qy      181 NEPEPELQRRGMKASRRQVWKRKADGGQDSGISSHFQWPMFHSIKHRYKSPPEESN 240
Db      181 NEPEPELQRRGMKASRRQVWKRKADGGQDSGISSHFQWPMFHSIKHRYKSPPEESN 240
Qy      241 QNGEGSYREAFETNSOVGLPIVYFSGRRPRLRLRPVLAETPEAFTEVAMVPEEGON 300
Db      241 QNGEGSYREAFETNSOVGLPIVYFSGRRPRLRLRPVLAETPEAFTEVAMVPEEGON 300
Qy      301 NPAILAGVFNCSHTVSDKGMALGIRSGDKGKRDAFFPSLCTDRYKATITLISHRYQ 360
Db      301 NPAILAGVFNCSHTVSDKGMALGIRSGDKGKRDAFFPSLCTDRYKATITLISHRYQ 360
Qy      361 PGWTHVAATYDGRHMAIYVDGTOVASLDSGSPINSPPMASSCSLLIGDSSSDGHYFR 420
Db      361 PGWTHVAATYDGRHMAIYVDGTOVASLDSGSPINSPPMASSCSLLIGDSSSDGHYFR 420
Qy      421 GHIGTLVFWSTALPOSHROHSSOHSSEEAETDVLTAASPPVNTVPPRDEKXPPLV 480
Db      421 GHIGTLVFWSTALPOSHROHSSOHSSEEAETDVLTAASPPVNTVPPRDEKXPPLV 480
Qy      481 LOGEPEPELISPLQPLCGQTVCDNVELISQYNGWYPLRGEKVIYQVNI CDDEGLNP 540
Db      481 LOGEPEPELISPLQPLCGQTVCDNVELISQYNGWYPLRGEKVIYQVNI CDDEGLNP 540
Qy      541 IYSEBOQIRLOHEALNEAFSRYNISWOLSVHOVNSTLRHRVTVNCEPSKIGNDHCDPEC 600
Db      541 IYSEBOQIRLOHEALNEAFSRYNISWOLSVHOVNSTLRHRVTVNCEPSKIGNDHCDPEC 600
Qy      601 EHPITGVGDGDCRLOGRCSYNNRRDGLCHVECNMMLNDFDGDCCDQVADVKTCTCPD 660
Db      601 EHPITGVGDGDCRLOGRCSYNNRRDGLCHVECNMMLNDFDGDCCDQVADVKTCTCPD 660
Qy      661 SPKRAVSVKELKEALQNSTHPLNTIYFASVREDLGAATPMDKDAVTHLGGIVLSPA 720
Db      661 SPKRAVSVKELKEALQNSTHPLNTIYFASVREDLGAATPMDKDAVTHLGGIVLSPA 720
Qy      721 YGMPGHTDTHIEVGHVGLYHVFKGVSERESCNDECKETVPSMETGDLCAADTAPTPKS 780
Db      721 YGMPGHTDTHIEVGHVGLYHVFKGVSERESCNDECKETVPSMETGDLCAADTAPTPKS 780
Qy      781 ELCEPEPEPTDTCGFTFPGAPFTNYSYTDNCTDNFTPOVARMHCYLDLYVQOMTES 840
Db      781 ELCEPEPEPTDTCGFTFPGAPFTNYSYTDNCTDNFTPOVARMHCYLDLYVQOMTES 840
Qy      841 RKPTPIPIPMVIGQTKSLTIHMLPISGVVYDRAAGSLCGATEBGTFRQVHTASSR 900
Db      841 RKPTPIPIPMVIGQTKSLTIHMLPISGVVYDRAAGSLCGATEBGTFRQVHTASSR 900
Qy      901 RVCDSGYWTEBEAVGPPVDVQCEPSIQAWSPEVHL YHMMNTVPCETEGCSLELFOHP 960
Db      901 RVCDSGYWTEBEAVGPPVDVQCEPSIQAWSPEVHL YHMMNTVPCETEGCSLELFOHP 960
Qy      961 VQADTLTLMTSFMESSQVLFDEILLENKESVHLGRLDFCIIPLTIKLAHDGKVS 1020
Db      961 VQADTLTLMTSFMESSQVLFDEILLENKESVHLGRLDFCIIPLTIKLAHDGKVS 1020
Qy      1021 KYVTFDERIEIDALLTSQSPSLCSGCRPVRYOVLRDPPASGLPVVYTHSHKFTDVE 1080
Db      1021 KYVTFDERIEIDALLTSQSPSLCSGCRPVRYOVLRDPPASGLPVVYTHSHKFTDVE 1080
Qy      1081 VTPGOMTOYOVLABAGELGASPLNHIHAPYCGGKVSERLGEBCDDGLVSGDGS 1140
Db      1081 VTPGOMTOYOVLABAGELGASPLNHIHAPYCGGKVSERLGEBCDDGLVSGDGS 1140

```

```

Qy      1141 KVCLEBEGFNCVGEPSLCYMYEAGDICEPERKTSIVDCGIYTPKGYLDQMATRAYSHE 1200
Db      1141 KVCLEBEGFNCVGEPSLCYMYEAGDICEPERKTSIVDCGIYTPKGYLDQMATRAYSHE 1200
Qy      1201 DKKCPVSLVTGEPHSLICTSYHPLDLPNHRPLTGWPPCVASENETODRSEOPGSLKE 1260
Db      1201 DKKCPVSLVTGEPHSLICTSYHPLDLPNHRPLTGWPPCVASENETODRSEOPGSLKE 1260
Qy      1261 DEWLVKCFNRPEBAIAIFLTTDGLVPEHQPVTYLVLTVRGNSHSLGTYSQCH 1320
Db      1261 DEWLVKCFNRPEBAIAIFLTTDGLVPEHQPVTYLVLTVRGNSHSLGTYSQCH 1320
Qy      1321 NPLIINTVHTQNVLFHHTTSVLTNFSPPRGISAVALTSSRIGLSAPSNCISEDEGON 1380
Db      1321 NPLIINTVHTQNVLFHHTTSVLTNFSPPRGISAVALTSSRIGLSAPSNCISEDEGON 1380
Qy      1381 QGOSCIHRPCGKODSCPSLLDHDADVNTCSIGPGLMKCAITQORFALQASSGOYIRPM 1440
Db      1381 QGOSCIHRPCGKODSCPSLLDHDADVNTCSIGPGLMKCAITQORFALQASSGOYIRPM 1440
Qy      1441 QKSIILTCSSGHWDOVNSCLPVDCGVPDPSLVNYANFSCSEGTFLKRCGISCVPAKQ 1500
Db      1441 QKSIILTCSSGHWDOVNSCLPVDCGVPDPSLVNYANFSCSEGTFLKRCGISCVPAKQ 1500
Qy      1501 GLSPMLTCLBEDGLMSLPEVYCKLECDAPRIILNANLLPHCLDNHVDGTICKYECRGY 1560
Db      1501 GLSPMLTCLBEDGLMSLPEVYCKLECDAPRIILNANLLPHCLDNHVDGTICKYECRGY 1560
Qy      1561 YVASAEGKVRNKLKIQCLEGGIWEQSGCIPVCEBPVPVFGWMECTNGFSLDSQCVL 1620
Db      1561 YVASAEGKVRNKLKIQCLEGGIWEQSGCIPVCEBPVPVFGWMECTNGFSLDSQCVL 1620
Qy      1621 NCNOERKPLILCTKGLMTQEFKLCENLOGECPPPSELSNVSVEYKCEGYGIGAVCSPL 1680
Db      1621 NCNOERKPLILCTKGLMTQEFKLCENLOGECPPPSELSNVSVEYKCEGYGIGAVCSPL 1680
Qy      1681 CVIPSPDPVMLPENITADTLEHMMEPVKVOSIYCTGRQHPDVLVHICIQSCPEFOA 1738
Db      1681 CVIPSPDPVMLPENITADTLEHMMEPVKVOSIYCTGRQHPDVLVHICIQSCPEFOA 1738

```

## RESULT 3

```

US-09-827-998-10
: Sequence 10, Application US/09827998
: Patent No. US20020102252A1
: GENERAL INFORMATION:
: APPLICANT: Gu, Yizhong
: TITLE OF INVENTION: NOVEL ISOFORMS OF HUMAN PREGNANCY-ASSOCIATED PROTEIN E
: FILE REFERENCE: MDHMRP-8
: CURRENT APPLICATION NUMBER: US/09/827, 998
: PRIOR FILING DATE: 2001-04-06
: PRIOR APPLICATION NUMBER: US 60/207,456
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: US 60/236,359
: NUMBER OF SEQ ID NOS: 1881
: SOFTWARE: Aecmla Sequence Listing Engine
: SEQ ID NO 10
: LENGTH: 1770
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-827-998-10

```

```

Query Match      85.5%; Score 1532; DB 10; Length 1770;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1732; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

Qy      1 MMCKILIRISLAILAGALCSANSELGWTTRKSLVEREHINOVILBGERCMLGAKVRRPR 60
Db      1 MMCKILIRISLAILAGALCSANSELGWTTRKSLVEREHINOVILBGERCMLGAKVRRPR 60

```

```

Qy 61 ASPOHLLFGVYPSRAGNYLRPYVGEBOEIHHTGRSKPDTEGNAVSLVPPDLTENPAGLRG 120
Db 61 ASPOHLLFGVYPSRAGNYLRPYVGEBOEIHHTGRSKPDTEGNAVSLVPPDLTENPAGLRG 120
Qy 121 AVEERPAAPWVDSP1GQSELLGDDDAYIGNORSKESLGEAGIQKGSAMAATTTTALIFTL 180
Db 121 AVEERPAAPWVDSP1GQSELLGDDDAYIGNORSKESLGEAGIQKGSAMAATTTTALIFTL 180
Qy 181 NEPEPEORRGMASROBOROYMKRRAEDQGDGSGISSHFOPMKSLKHKRKSPPREESN 240
Db 181 NEPEPEORRGMASROBOROYMKRRAEDQGDGSGISSHFOPMKSLKHKRKSPPREESN 240
Qy 241 ONGEGSYREAEATNSQVGLPILVFSGRERLLRPEVLAIIPREAFTEAMVPEEGON 300
Db 241 ONGEGSYREAEATNSQVGLPILVFSGRERLLRPEVLAIIPREAFTEAMVPEEGON 300
Qy 301 NPATIAGVFDCSHTVSDKMGALGRSGDKDKRDARFFSLCTDRVKKATILLSHSRVQ 360
Db 301 NPATIAGVFDCSHTVSDKMGALGRSGDKDKRDARFFSLCTDRVKKATILLSHSRVQ 360
Qy 361 PGTHVAATYDGRHMLYVDGTOVASLDOSGPLNSPFMASCRSLLDGDSSEDEGHYR 420
Db 361 PGTHVAATYDGRHMLYVDGTOVASLDOSGPLNSPFMASCRSLLDGDSSEDEGHYR 420
Qy 421 GHLGLVFWSTALPQSHFOHSSSGREERATDLVLTASFEPTVTEWVPFDEKXPRLV 480
Db 421 GHLGLVFWSTALPQSHFOHSSSGREERATDLVLTASFEPTVTEWVPFDEKXPRLV 480
Qy 481 LOGEPEPEELISPLQPLCGQTCVCDNVELISQYNGWPLREKXYIRYOYVNI CDDEGLN 540
Db 481 LOGEPEPEELISPLQPLCGQTCVCDNVELISQYNGWPLREKXYIRYOYVNI CDDEGLN 540
Qy 541 IVSEBOIRLOHEALNEAFSRNISMOLSVHQNSTLRHRVVLVNCESKTI GNHCDDEC 600
Db 541 IVSEBOIRLOHEALNEAFSRNISMOLSVHQNSTLRHRVVLVNCESKTI GNHCDDEC 600
Qy 601 EHPILGYDGDGCRLOGRCYSWNRBDGLCHVECNMNLNFDGDCDDPQVADVRKTCFDPD 660
Db 601 EHPILGYDGDGCRLOGRCYSWNRBDGLCHVECNMNLNFDGDCDDPQVADVRKTCFDPD 660
Qy 661 SPKAAVSVKELKELQUNSTHPLNTYFASVREDLAGAATWPKDAVTHLGGIVLSPA 720
Db 661 SPKAAVSVKELKELQUNSTHPLNTYFASVREDLAGAATWPKDAVTHLGGIVLSPA 720
Qy 721 YYGMPGHTDTHIEGHVGLGLYHVFKGVSERSSCNDPCKEYVPSMETGDLADTAAPRKS 780
Db 721 YYGMPGHTDTHIEGHVGLGLYHVFKGVSERSSCNDPCKEYVPSMETGDLADTAAPRKS 780
Qy 781 ELCEPEPEPTSDTCGFTFRPGAPFTNYMSYTDNCTDNFTPNQVARMHCYLDLVYQOMTES 840
Db 781 ELCEPEPEPTSDTCGFTFRPGAPFTNYMSYTDNCTDNFTPNQVARMHCYLDLVYQOMTES 840
Qy 841 RKPTPIPIPPVNTIGQTNKSLTIHMLPRTISGVYVDRASGLCGACTEDGTFRQYHTASSR 900
Db 841 RKPTPIPIPPVNTIGQTNKSLTIHMLPRTISGVYVDRASGLCGACTEDGTFRQYHTASSR 900
Qy 901 RVCSGSGWTEBBAVGPDPVOPCEPSLOAMSPEVHLVHMNTVPCPEGCSLELTFQHP 960
Db 901 RVCSGSGWTEBBAVGPDPVOPCEPSLOAMSPEVHLVHMNTVPCPEGCSLELTFQHP 960
Qy 961 VQADTLTLMTVSFFESSQVLFDTIELLENKESVHLGLDTFCDIPLTIKHLVHGKVSGLV 1020
Db 961 VQADTLTLMTVSFFESSQVLFDTIELLENKESVHLGLDTFCDIPLTIKHLVHGKVSGLV 1020
Qy 1021 KYVTFDERIEIDALLISQHPSPICSGCRPVRYOVLADPPASGLPVVVTSHRKTDFVE 1080
Db 1021 KYVTFDERIEIDALLISQHPSPICSGCRPVRYOVLADPPASGLPVVVTSHRKTDFVE 1080
Qy 1081 VTPOGMVOYVLAABGELGASPLNTHIGAPYCGDQKVSBRGEBECCDDDLVSGDSCS 1140
Db 1081 VTPOGMVOYVLAABGELGASPLNTHIGAPYCGDQKVSBRGEBECCDDDLVSGDSCS 1140
Qy 1141 KVCELEBEGFNCVGEPSLICYMEGDGICPEPFERKTSIVDCGIYTPKGYLDQWATRAYSSHE 1200

```

```

Db 1141 KVCELEBEGFNCVGEPSLICYMEGDGICPEPFERKTSIVDCGIYTPKGYLDQWATRAYSSHE 1200
Qy 1201 DKKKCPVSLVTGEPHSLICTSYHNDLBNHRPLTGWPFVAVS ENEOTDDREOEGSLKKE 1260
Db 1201 DKKKCPVSLVTGEPHSLICTSYHNDLBNHRPLTGWPFVAVS ENEOTDDREOEGSLKKE 1260
Qy 1261 DEWLKVCNRPGEARAFIFLFTDGLVPEHQQPTVTLVTVRGSNHS LGTYGSCQH 1320
Db 1261 DEWLKVCNRPGEARAFIFLFTDGLVPEHQQPTVTLVTVRGSNHS LGTYGSCQH 1320
Qy 1321 NPLIINTVHHQNYLFHHTTSVLLNFSSPRVIGISAVALTSSRIIGLSAPNSCISEDEGQNH 1380
Db 1321 NPLIINTVHHQNYLFHHTTSVLLNFSSPRVIGISAVALTSSRIIGLSAPNSCISEDEGQNH 1380
Qy 1381 OGOSCHIRPGCKDSCPSLILDHADVNCSTISPGMLKCAITQORGFALQASSGOYIRPM 1440
Db 1381 OGOSCHIRPGCKDSCPSLILDHADVNCSTISPGMLKCAITQORGFALQASSGOYIRPM 1440
Qy 1441 QKEILLTCSGHNDOVNSCLPVDGVPDPSLVVYANFSCSEGTKFLKRCISICVPAKLQ 1500
Db 1441 QKEILLTCSGHNDOVNSCLPVDGVPDPSLVVYANFSCSEGTKFLKRCISICVPAKLQ 1500
Qy 1501 GLSPWILTCLEBDGLMSLPEVYCKLECDAPRIILNANILLPHCLQDNHDVGTICKYECKPGY 1560
Db 1501 GLSPWILTCLEBDGLMSLPEVYCKLECDAPRIILNANILLPHCLQDNHDVGTICKYECKPGY 1560
Qy 1561 YVASASAGKRNKLLKIQCEGGIWEQSGCLPVVCEPFPVFEGMVECTNGFSLDSCVL 1620
Db 1561 YVASASAGKRNKLLKIQCEGGIWEQSGCLPVVCEPFPVFEGMVECTNGFSLDSCVL 1620
Qy 1621 NCNOREKLPILCTKEGLMTQEFPLCENLQGECPPPSELSVYKCEQSGYIGAVCSPL 1680
Db 1621 NCNOREKLPILCTKEGLMTQEFPLCENLQGECPPPSELSVYKCEQSGYIGAVCSPL 1680
Qy 1681 CVIPSDPVMLENITADTLEHMMEPVKVQSVCTGRQNHPPVLVHCTIOSCE 1734
Db 1681 CVIPSDPVMLENITADTLEHMMEPVKVQSVCTGRQNHPPVLVHCTIOSCE 1734

RESULT 4
US-09-827-998-16
; Sequence 16, Application US/09827998
; Patent No. US20020102252A1
; GENERAL INFORMATION:
; APPLICANT: Gu, Yizhong
; TITLE OF INVENTION: NOVEL ISOFORMS OF HUMAN PREGNANCY-ASSOCIATED PROTEIN E
; FILE REFERENCE: MDHMOF-8
; CURRENT APPLICATION NUMBER: US/09/827,998
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 1881
; SOFTWARE: Aeonica Sequence Listing Engine
; SEQ ID NO 16
; LENGTH: 1385
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-827-998-16

Query Match 32.0%; Score 574; DB 10; Length 1385;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 1074; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 713 GGATLSPAYYGMPTDTHIEGHVGLGLYHVFKGVSERSSCNDPCKEYVPSMETGDLCA 772
Db 307 GGATLSPAYYGMPTDTHIEGHVGLGLYHVFKGVSERSSCNDPCKEYVPSMETGDLCA 366
Qy 773 DTAPTKSELCREPEPTSDTCGFTFRPGAPFTNYMSYTDNCTDNFTPNQVARMHCYLDL 832

```



```

Db      367  DTAPPKSELCREBEPTSDTCGTRFGAPFTNMYSTDNCTDNTPNQVARNHCYLDL 426
Qy      833  VVOQMTESRKPTPIPIPMVIGOTNKSJLTJHMLPPIGYYVYDRASGLACACTEDGFRQ 892
Db      427  VVOQMTESRKPTPIPIPMVIGOTNKSJLTJHMLPPIGYYVYDRASGLACACTEDGFRQ 486
Qy      893  YHTASSRRVCDSSGYVTPEEAVGPDPVDPCEBSLQAMSEVHLVYMMNTVPCPTGCS 952
Db      487  YHTASSRRVCDSSGYVTPEEAVGPDPVDPCEBSLQAMSEVHLVYMMNTVPCPTGCS 546
Qy      953  LELLFOHPVADTLTMTVTSFPMESSQVLTETELLENKSSVHLGELDTCDIPLTKH 1012
Db      547  LELLFOHPVADTLTMTVTSFPMESSQVLTETELLENKSSVHLGELDTCDIPLTKH 606
Qy      1013  VDKVSGVKKYTFDERIEIDAALLTSOPHSPLCSGCRPVYQVLRDPPASGLPVVTHS 1072
Db      607  VDKVSGVKKYTFDERIEIDAALLTSOPHSPLCSGCRPVYQVLRDPPASGLPVVTHS 666
Qy      1073  HRKFTDVEVTPEGMYQVYLAAGELGEASPLNHIHGAAPYCGDGKYSRLGECDDGD 1132
Db      667  HRKFTDVEVTPEGMYQVYLAAGELGEASPLNHIHGAAPYCGDGKYSRLGECDDGD 726
Qy      1133  LVSGDGSKVCLEBEGNCVCEBPLCYMREGDGTCEPEKRTSIVDCGITTCKYLDQMA 1192
Db      727  LVSGDGSKVCLEBEGNCVCEBPLCYMREGDGTCEPEKRTSIVDCGITTCKYLDQMA 786
Qy      1193  TRAYSSHEDKCKKCVSLVTGEPHSLICTSYHPLDPLNRPPLTGMPCVASENETODDSEQ 1252
Db      787  TRAYSSHEDKCKKCVSLVTGEPHSLICTSYHPLDPLNRPPLTGMPCVASENETODDSEQ 846
Qy      1253  PEGSLKKEDEVMVKVCFNRGGEARAIPIFLTTDGLVGEHQOFTVTLVLTDRVSNHSLG 1312
Db      847  PEGSLKKEDEVMVKVCFNRGGEARAIPIFLTTDGLVGEHQOFTVTLVLTDRVSNHSLG 906
Qy      1313  TYGLSCCHNPLIINVTHQNVLFHTTSVTLNSSFPRVGSIAVALRTSSIGLSAPENCI 1372
Db      907  TYGLSCCHNPLIINVTHQNVLFHTTSVTLNSSFPRVGSIAVALRTSSIGLSAPENCI 966
Qy      1373  SEDEGONHOGOSCIHRPCGKODSCPSLLDHDVNTSISGPGIMKCAITCOGFALQAS 1432
Db      967  SEDEGONHOGOSCIHRPCGKODSCPSLLDHDVNTSISGPGIMKCAITCOGFALQAS 1026
Qy      1433  SGOYIRPMQKEIILTCSSGHMDQVNSCLPYDCGVPDLSLVNYANFSCSEGTFFLRCSIS 1492
Db      1027  SGOYIRPMQKEIILTCSSGHMDQVNSCLPYDCGVPDLSLVNYANFSCSEGTFFLRCSIS 1086
Qy      1493  CVPPAKIQLSPMLTCLJEDGIMSLPEVYCKLECDAPPIIANMLLPHCLQDNHVDGTIC 1552
Db      1087  CVPPAKIQLSPMLTCLJEDGIMSLPEVYCKLECDAPPIIANMLLPHCLQDNHVDGTIC 1146
Qy      1553  KYEGKPGYVYAESAGKVRNKLKIOCLEGIWEGSCIPVVCBPPVPEBGMVECTNGF 1612
Db      1147  KYEGKPGYVYAESAGKVRNKLKIOCLEGIWEGSCIPVVCBPPVPEBGMVECTNGF 1206
Qy      1613  SLDSQCVLANQOEREKLPILCTKEGLWTOEFKLCENIQGECPPPSSELANVEYKCEGYG 1672
Db      1207  SLDSQCVLANQOEREKLPILCTKEGLWTOEFKLCENIQGECPPPSSELANVEYKCEGYG 1266
Qy      1673  IGAVCSPLCVIPSPDPVMLPENITADTLJHMEBVKVQSVICTGRARQWHDHPVLVHCIOG 1732
Db      1267  IGAVCSPLCVIPSPDPVMLPENITADTLJHMEBVKVQSVICTGRARQWHDHPVLVHCIOG 1326
Qy      1733  CEFPQADGMDCTINNRAYCHYDGDCCSSTLSSKVIIPFADCDLDBCTCRDPAEENO 1791
Db      1327  CEFPQADGMDCTINNRAYCHYDGDCCSSTLSSKVIIPFADCDLDBCTCRDPAEENO 1385

```

RESULT 5  
 US-09-864-761-34265  
 ; Sequence 34265 Application US/09864761  
 ; Patent No. US20020048763A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Penn, Sharon G.

```

; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 09/864,761
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 4917
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 34265
; LENGTH: 192
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL031734.9
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.4
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.6
; OTHER INFORMATION: EST HUMAN HIT: A0140701.1, EVALUATE 2.00e-53
; OTHER INFORMATION: SWISSPROT HIT: P07207, EVALUATE 3.00e-04
; US-09-864-761-34265

```

Query Match 10.7%; Score 192; DB 9; Length 192;  
 Best Local Similarity 100.0%; Pred. No. 7.2e-180;  
 Matches 192; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy      472  DEKYPRLVQLGEPPELISPLQGPLCGQTVCDNVELISQYNGYPLRGEKVIYRQVNN 531
Db      1    DEKYPRLVQLGEPPELISPLQGPLCGQTVCDNVELISQYNGYPLRGEKVIYRQVNN 60

```

QY 532 ICDEGLNPIYSEBQIRLOHEALNEAFSRYNISWQLSVHQVHNSTLRRVVLVNCESPSKI 591  
DB 61 ICDEGLNPIYSEBQIRLOHEALNEAFSRYNISWQLSVHQVHNSTLRRVVLVNCESPSKI 120  
QY 592 GNDHCDPECEHPLTGYDGGDGLGRCYSNNRBDGLCHVECNMNLNPDGDCDDCPVAD 651  
DB 121 GNDHCDPECEHPLTGYDGGDGLGRCYSNNRBDGLCHVECNMNLNPDGDCDDCPVAD 180  
QY 652 VRKTCFDPDSPK 663  
DB 181 VRKTCFDPDSPK 192

RESULT 6  
US-09-864-761-34264  
Sequence 34264, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: Aecomica-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
PRIOR FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263,6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 34264  
LENGTH: 70  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AL031734.9  
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.6

OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.3  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 44  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.6  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.4  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.9  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.6  
OTHER INFORMATION: SWISSPROT HIT: P28977, EVALUATE 1.60e+00  
OTHER INFORMATION: EST\_HUMAN HIT: BF366974.1, EVALUATE 4.00e-36  
US-09-864-761-34264

Query Match 3.9%; Score 70; DB 9; Length 70;  
Best Local Similarity 100.0%; Pred. No. 2.5e-60;  
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 308 VFDNCSHTTSDKMGALGIRSGKDKGRDAFFFSLCTDRYKATILISHSRYPGTWTHV 367  
DB 1 VFDNCSHTTSDKMGALGIRSGKDKGRDAFFFSLCTDRYKATILISHSRYPGTWTHV 60  
QY 368 AATYDGRHMA 377  
DB 61 AATYDGRHMA 70

RESULT 7  
US-09-864-761-34262  
Sequence 34262, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: Aecomica-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
PRIOR FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263,6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408

PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Annotmax Sequence Listing Engine vers. 1.1  
SEQ ID NO 34262  
LENGTH: 63  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AL031734.9  
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 0.87  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.84  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.82  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.89  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.89  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.8  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.79  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1  
OTHER INFORMATION: EST\_HUMAN HIT: A1075970.1, EVALUE 7.00e-03  
US-09-864-76134262

Query Match 3.5%; Score 63; DB 9; Length 63;  
Best Local Similarity 100.0%; Pred. No. 1.7e-53;  
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 244 GEGSYRAETFNQVGLPIYFGRRRLRLRPVLAETREAVTAWYKPEGGONPA 303

DB 1 GEGSYRAETFNQVGLPIYFGRRRLRLRPVLAETREAVTAWYKPEGGONPA 60

QY 304 IIA 306  
DB 61 IIA 63

RESULT 8  
US-09-827-998-7  
Sequence 7, Application US/09827998  
Patent No. US20020102252A1  
GENERAL INFORMATION:  
APPLICANT: Shannon, Mark  
TITLE OF INVENTION: NOVEL ISOFORMS OF HUMAN PREGNANCY-ASSOCIATED PROTEIN E  
FILE REFERENCE: MDHMRP-8  
CURRENT APPLICATION NUMBER: US/09/827,998  
CURRENT FILING DATE: 2001-04-06  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
NUMBER OF SEQ ID NOS: 1881  
SOFTWARE: Aeonica Sequence Listing Engine  
SEQ ID NO 7  
LENGTH: 19  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-827-998-7

Query Match 1.1%; Score 19; DB 10; Length 19;  
Best Local Similarity 100.0%; Pred. No. 1e-10;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MMCKIIRISLAIAGNAL 19  
DB 1 MMCKIIRISLAIAGNAL 19

RESULT 9  
US-09-983-025-25  
Sequence 25, Application US/09983025  
Publication No. US20030124529A1

GENERAL INFORMATION:  
APPLICANT: OXVIG, Claus  
TITLE OF INVENTION: PREGNANCY-ASSOCIATED PLASMA PROTEIN-A2 (PAPP-A2)  
FILE REFERENCE: OXVIG-1A  
CURRENT APPLICATION NUMBER: US/09/983,025  
CURRENT FILING DATE: 2001-10-22  
PRIOR APPLICATION NUMBER: US 60/241,840  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: DK PA 2000 01571  
PRIOR FILING DATE: 2000-10-20  
NUMBER OF SEQ ID NOS: 25  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 25  
LENGTH: 1627  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-983-025-25

Query Match 1.0%; Score 18; DB 11; Length 1627;  
Best Local Similarity 100.0%; Pred. No. 5.5e-08;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 818 FTNPQVARMHCYLDLVYQ 835  
DB 647 FTNPQVARMHCYLDLVYQ 664

RESULT 10  
US-10-295-027-663  
Sequence 663, Application US/10295027  
Publication No. US2003022350A1  
GENERAL INFORMATION:  
APPLICANT: Afar, Daniel  
APPLICANT: Aziz, Natasha  
APPLICANT: Ginsberg, Wendy M.  
APPLICANT: Gish, Kurt C.  
APPLICANT: Glynn, Richard  
APPLICANT: Hevez, Peter A.  
APPLICANT: Mack, David H.  
APPLICANT: Murray, Richard  
APPLICANT: Watson, Susan R.  
TITLE OF INVENTION: Methods of diagnosis of Cancer. Compositions and  
TITLE OF INVENTION: Methods of Screening for Modulators of Cancer  
FILE REFERENCE: 018501-01250005  
CURRENT APPLICATION NUMBER: US/10/295,027  
CURRENT FILING DATE: 2002-11-13  
PRIOR APPLICATION NUMBER: US 09/663,733  
PRIOR FILING DATE: 2000-09-15  
PRIOR APPLICATION NUMBER: US 60/350,666  
PRIOR FILING DATE: 2001-11-13  
PRIOR APPLICATION NUMBER: US 60/335,394  
PRIOR FILING DATE: 2001-11-15  
PRIOR APPLICATION NUMBER: US 60/332,464  
PRIOR FILING DATE: 2001-11-21  
PRIOR APPLICATION NUMBER: US 60/334,393  
PRIOR FILING DATE: 2001-11-29  
PRIOR APPLICATION NUMBER: US 60/340,376  
PRIOR FILING DATE: 2001-12-14  
PRIOR APPLICATION NUMBER: US 60/347,211  
PRIOR FILING DATE: 2002-01-08  
PRIOR APPLICATION NUMBER: US 60/347,349  
PRIOR FILING DATE: 2002-01-10  
PRIOR APPLICATION NUMBER: US 60/355,250  
PRIOR FILING DATE: 2002-02-08  
PRIOR APPLICATION NUMBER: US 60/356,714  
PRIOR FILING DATE: 2002-02-13  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 1386  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 663  
LENGTH: 1627

TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-295-027-663

Query Match 1.0%; Score 18; DB 12; Length 1627;  
Best Local Similarity 100.0%; Pred. No. 5.5e-08;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 818 FTPOVARMHCYLDLVYQ 835  
DB 647 FTPOVARMHCYLDLVYQ 664

RESULT 11  
US-09-827-998-18  
Sequence 18, Application US/09827998  
Patent No. US20020102252A1  
GENERAL INFORMATION:  
APPLICANT: Shannon, Mark  
TITLE OF INVENTION: NOVEL ISOFORMS OF HUMAN PREGNANCY-ASSOCIATED PROTEIN B  
FILE REFERENCE: MPMORF-8  
CURRENT APPLICATION NUMBER: US/09/827,998  
PRIOR FILING DATE: 2001-04-06  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
NUMBER OF SEQ ID NOS: 1881  
SOFTWARE: Aeomica Sequence Listing Engine  
SEQ ID NO 18  
LENGTH: 20  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-827-998-18

Query Match 0.6%; Score 11; DB 10; Length 20;  
Best Local Similarity 100.0%; Pred. No. 0.008;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 713 GGIVLSPAYYG 723  
DB 10 GGIVLSPAYYG 20

RESULT 12  
US-10-369-493-18401  
Sequence 18401, Application US/10369493  
Publication No. US20030233675A1  
GENERAL INFORMATION:  
APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xiaofeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
FILE REFERENCE: 38-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369,493  
PRIOR FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360,039  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 18401  
LENGTH: 502  
TYPE: PRT  
ORGANISM: Lactococcus lactis  
US-10-369-493-18401

Query Match 0.5%; Score 9; DB 12; Length 502;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1014 DGRVGVKV 1022  
DB 217 DGRVGVKV 225

RESULT 13  
US-10-115-072-7  
Sequence 7, Application US/10115072  
Publication No. US20030105003A1  
GENERAL INFORMATION:  
APPLICANT: NILSSON, JAN  
APPLICANT: SHAH, FREDIMAN K.  
TITLE OF INVENTION: PEPTIDE-BASED IMMUNIZATION THERAPY FOR TREATMENT OF  
TITLE OF INVENTION: ATHEROSCLEROSIS AND DEVELOPMENT OF PEPTIDE-BASED ASSAY  
TITLE OF INVENTION: FOR DETERMINATION OF IMMUNE RESPONSES AGAINST OXIDIZED  
TITLE OF INVENTION: LOW DENSITY LIPOPROTEIN  
FILE REFERENCE: 03940.0057  
CURRENT APPLICATION NUMBER: US/10/115,072  
CURRENT FILING DATE: 2002-08-20  
PRIOR APPLICATION NUMBER: 60/281,410  
PRIOR FILING DATE: 2001-05-04  
PRIOR APPLICATION NUMBER: SE 0101232-7  
PRIOR FILING DATE: 2001-05-04  
PRIOR APPLICATION NUMBER: SE 0103754-8  
PRIOR FILING DATE: 2001-09-11  
NUMBER OF SEQ ID NOS: 38  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 7  
LENGTH: 20  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
OTHER INFORMATION: peptide  
US-10-115-072-7

Query Match 0.4%; Score 8; DB 15; Length 20;  
Best Local Similarity 100.0%; Pred. No. 7.1;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1312 GTYGLSCQ 1319  
DB 2 GTYGLSCQ 9

RESULT 14  
US-10-029-386-30575  
Sequence 30575, Application US/10029386  
Publication No. US20030194704A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G  
FILE REFERENCE: AEOMICA-X-2  
CURRENT APPLICATION NUMBER: US/10/029,386  
CURRENT FILING DATE: 2001-12-20  
NUMBER OF SEQ ID NOS: 34288  
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 30575  
LENGTH: 28  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC011799.5  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.37  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.4  
US-10-029-386-30575

Query Match 0.4%; Score 8; DB 12; Length 28;  
Best Local Similarity 100.0%; Pred. No. 9.7;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1760 SSTLSSKK 1767  
Db 15 SSTLSSKK 22

RESULT 15  
US-10-078-929-84  
; Sequence 84, Application US/10078929  
; Publication No. US20020152497A1  
; GENERAL INFORMATION:  
; APPLICANT: Rafalski, Antoni  
; APPLICANT: Miao, Guo-Hua  
; APPLICANT: Falco, Saverio Carl  
; APPLICANT: Sakai, Hajime  
; APPLICANT: Famodu, Omolayo O.  
; APPLICANT: Ogell, Joan T.  
; APPLICANT: Meyers, Blake  
; APPLICANT: Thorpe, Catherine  
; APPLICANT: Weng, Zude  
; TITLE OF INVENTION: Nucleic Acid Fragments Encoding Proteins Involved in  
; TITLE OF INVENTION: Stress Response  
; FILE REFERENCE: BR1357 US NA  
; CURRENT APPLICATION NUMBER: US/10/078,929  
; PRIOR FILING DATE: 2002-02-19  
; PRIOR APPLICATION NUMBER: 09/566,394  
; PRIOR FILING DATE: 2000-05-05  
; PRIOR APPLICATION NUMBER: 60/133038  
; PRIOR FILING DATE: 1999-05-07  
; PRIOR APPLICATION NUMBER: 60/133042  
; PRIOR FILING DATE: 1999-05-07  
; PRIOR APPLICATION NUMBER: 60/133427  
; PRIOR FILING DATE: 1999-05-11  
; PRIOR APPLICATION NUMBER: 60/133437  
; PRIOR FILING DATE: 1999-05-11  
; PRIOR APPLICATION NUMBER: 60/133428  
; PRIOR FILING DATE: 1999-05-11  
; PRIOR APPLICATION NUMBER: 60/133438  
; PRIOR FILING DATE: 1999-05-11  
; PRIOR APPLICATION NUMBER: 60/133436  
; PRIOR FILING DATE: 1999-05-11  
; PRIOR APPLICATION NUMBER: 60/137667  
; PRIOR FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 208  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 84  
; LENGTH: 164  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; NAME/KEY: UNSURE  
; LOCATION: (22)  
; NAME/KEY: UNSURE  
; LOCATION: (50)  
; NAME/KEY: UNSURE  
; LOCATION: (56)  
; NAME/KEY: UNSURE  
; LOCATION: (59)  
; NAME/KEY: UNSURE  
; LOCATION: (99)  
; NAME/KEY: UNSURE  
; LOCATION: (121)  
; NAME/KEY: UNSURE  
; LOCATION: (140)  
; NAME/KEY: UNSURE  
; LOCATION: (150)  
; NAME/KEY: UNSURE  
; LOCATION: (158)  
; NAME/KEY: UNSURE  
; LOCATION: (163)  
US-10-078-929-84

Query Match

0.4%; Score 8; DB 14; Length 164;

Best Local Similarity 100.0%; Pred. No. 47;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 167 AMAATTTT 174  
Db 13 AMAATTTT 20

Search completed: January 2, 2004, 16:21:19  
Job time : 52 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2004, 16:02:37 ; Search time 28 Seconds

(without alignments)  
6151.361 Million cell updates/sec

Title: US-09-983-025a-2

Sequence: 1 MWCLKILRLSLATLACGALC.....AADCDDECTCRDPKAEHQ 1791

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

1: PIR\_76:\*  
2: PIR1:\*  
3: PIR2:\*  
4: PIR3:\*  
5: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3916.5	39.7	1627	2 S65464	pregnancy-associated complement C3b/C4b
2	292.5	3.0	2489	2 I73012	complement C3b/C4b
3	289	2.9	2014	2 I36936	MEGF6 protein - ra
4	282.5	2.9	1574	2 T13954	P-selectin precurs
5	249	2.5	830	2 A30359	E-selectin precurs
6	247	2.5	612	2 B42755	E-selectin precurs
7	244	2.5	1091	1 P10009	complement C3d/Bps
8	238.5	2.4	1234	1 NBM5H	complement factor
9	230.5	2.3	610	2 A35066	E-selectin precurs
10	228	2.3	551	2 I46709	endothelial leukoc
11	227.5	2.3	768	2 I53821	P-selectin - rat
12	225	2.3	1394	2 A35626	transforming growt
13	224	2.3	1620	2 T27283	hypothetical prote
14	219	2.2	768	2 A42755	P-selectin precurs
15	218.5	2.2	597	1 S53711	C4BP alpha chain p
16	218.5	2.2	1025	1 A43526	complement C3d/Bps
17	218	2.2	1712	2 A38261	masking protein pr
18	217	2.2	5376	2 T42215	zonadhesin - mouse
19	215	2.2	381	1 B26359	decay-accelerating
20	215	2.2	440	2 A26359	decay-accelerating
21	213.5	2.1	1231	1 NBM5H	complement factor
22	203	2.0	2471	2 A49128	cell-fate determin
23	202	2.0	646	2 UN0473	P-selectin precurs
24	200	2.0	558	2 S57953	C4BP protein alpha
25	200	2.0	3133	2 S52093	hemocytin - silkw
26	198	2.0	2437	2 S42612	transmembrane prot
27	197.5	2.0	610	1 I46001	C4b-binding protei
28	196.5	2.0	1820	2 A55494	latent transformin
29	196	2.0	485	2 S36772	E-selectin - bovin

30	196	2.0	1548	2 S34583	serine proteinase
31	196	2.0	2150	2 T32497	hypothetical prote
32	193.5	2.0	597	1 NBM5H	C4b-binding protei
33	193.5	2.0	2871	2 A55624	fibrinogen-1 precu
34	193	2.0	668	2 A46013	coagulation factor
35	192.5	2.0	2555	2 A46013	decay-accelerating
36	191.5	1.9	340	2 S56551	factor H - bovine
37	189	1.9	669	2 S56551	factor H - bovine
38	189	1.9	1429	2 S06434	homeotic protein 1
39	189	1.9	3084	1 MMSA	laminin alpha-1 ch
40	188.5	1.9	808	2 D35069	complement factor
41	188.5	1.9	2871	2 A55567	fibrinogen-1 - bovi
42	188.5	1.9	3051	2 S42373	hypothetical prote
43	188	1.9	1203	2 A49175	Notch B protein -
44	187.5	1.9	1111	2 T26972	hypothetical prote
45	187	1.9	482	2 JC5092	E-selectin - pig

#### ALIGNMENTS

RESULT 1  
S65464  
pregnancy-associated plasma protein A precursor - human  
N:Alternate names: PAP-A  
C:Species: Homo sapiens (man)  
C>Date: 22-Nov-1996 #sequence revision 22-Nov-1996 #text\_change 05-Nov-1999  
C:Accession: S65464; S65463; A54220; I38097  
R:Haaning, J.; Oxvig, C.; Overgaard, M.T.; Ebbesen, P.; Kristensen, T.; Sottrup-Jensen, J.  
submitted to the EMBL Data Library, June 1995  
A:Description: Complete cDNA sequence of the preproform of human pregnancy-associated p1.  
A:Reference number: S65464  
A:Accession: S65464  
A:Molecule type: mRNA  
A:Residues: 1-1627 <HAA>  
A:Cross-references: EMBL:U28727; NID:G1142969; PIDN:AAC50543.1; PID:G1142970  
R:Haaning, J.; Oxvig, C.; Overgaard, M.T.; Ebbesen, P.; Kristensen, T.; Sottrup-Jensen, J.  
Eur. J. Biochem. 237, 159-163, 1996  
A:Title: Complete cDNA sequence of the preproform of human pregnancy-associated plasma p.  
A:Reference number: S65463; MUID:96203921; PMID:8620868  
A:Molecule type: mRNA  
A:Accession: S65463  
A:Molecule type: mRNA  
A:Residues: 1-102 <HAM>  
A:Cross-references: EMBL:U28727  
A>Note: the authors translated the codon CGA for residue 101 as Thr  
R:Kristensen, T.; Oxvig, C.; Sand, O.; Moller, N.P.H.; Sottrup-Jensen, L.  
Biochemistry 33, 1592-1598, 1994  
A:Title: Amino acid sequence of human pregnancy-associated plasma protein-A derived from  
A:Reference number: A54220; MUID:94146014; PMID:7508748  
A:Accession: A54220  
A:Molecule type: mRNA  
A:Residues: 77-1627 <KRI>  
A:Cross-references: GB:X68280; NID:G394649; PIDN:CAA48341.1; PID:G394650  
R:Oxvig, C.; Sand, O.; Kristensen, T.; Gleich, G.J.; Sottrup-Jensen, L.  
J. Biol. Chem. 269, 12243-12246, 1993  
A:Title: Circulating human pregnancy-associated plasma protein-A is disulfide-bridged to  
A:Reference number: I38097; MUID:93286045; PMID:7685339  
A:Accession: I38097  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 77-1627 <RES>  
A:Cross-references: EMBL:X68280; NID:G394649; PIDN:CAA48341.1; PID:G394650  
C:Genetics:  
A:Gene: GDB:PAPPA  
A:Cross-references: GDB:134729; OMIM:176385  
A:Map position: 9q33.1-9q33.1  
F:1-22/Domain: signal sequence #status predicted <SIG>  
F:23-80/Domain: propeptide #status predicted <PRO>  
F:81-1627/Product: pregnancy-associated plasma protein A #status predicted <MAT>  
Query Match 39.7%; Score 3916.5; DB 2; Length 1627;  
Best Local Similarity 45.8%; Pred. No. 1.86-246;  
Matches 718; Conservative 296; Mismatches 499; Indels 53; Gaps 25;

QY	249	REAFTFNSOVL-P--LTFESGRRRRL--LREVLAEIPREAFTEAMKPEGGONPALLI	305
Dd	80	REAGAEETSEBPSPBALTFPSRGEOLRVLRDL--ELPRDPAITQWLRAGGQSPAVI	137
QY	306	AGEVDCNSHTYSDKGMALGIRSGDKGRDARFPFSLCTDRKVKATILLISHRYQPGWT	365
Dd	138	TGLYDKCSYIIRSDRGWVVGHTITSDQDKDPRYFSSLKTDBARQYTTINARSLPGQWV	197
QY	366	HVAATYQGRHMLVVDGTQVASSLDDQSPPLNSPFMAORSLLDGDSESDHRYRHLGT	425
Dd	198	YLAAITYDQFEMKLYVNGAQVATSGEQVGITSPITQCKVLMUG--SALNHNRYGIEH	255
QY	426	LVFMASTALPOSHFOHSSGHSGBEATDLYTASPEPNTWEPVFRDEKYPRLEW--LOG	483
Dd	256	FSLMKVARTQGBELISDMETHGAHTALPOLLQEMWMDVYKHAMSPMKDSSPKVEFSNMG	315
QY	484	FEPEBELISPLQPLRCQGTCDVNDWELISQNGYMPLEGEKIRYQVYNTCDDEGLNPVS	543
Dd	316	FLLO-----TSLDEPPLCQGTQLCDNTEVLASYNQLSFRQPKVVRVYVLLYEDDHKNPVYT	371
QY	544	EOQIRLOHEALNEAFSRYNISMOTSVOHVNSTLRHVLVNCPESKIGNHCPEGCHP	603
Dd	372	REQYDFQHOLAENAKQYNTISWELDVENSLSLRRIILANCDLSKIDGENDCEPECHT	431
QY	604	LTYGDGDCR--LQGRCSYWNRRDGLCHVECNMMLNDFDGDCCDPQVADVAKTCFDPDP	662
Dd	432	LTHGDGGCRHLRHPAFVKOHNGCDMDCYVERFNPFGGECDCDPEITNTVYOTCFDPDSP	491
QY	663	KRAYMSVELKEALQNTSHFLNITYASVVEDLAGATWMDDAVTHLGGVILSPAYY	722
Dd	492	HRAVYDVELKNLILLDGSTHNLNIFFASSSEELAGVATWMDKEALMHLGIVLNPSEY	551
QY	723	GMPGHTDMIEVGVHVLGVYHFKVGSERESCNDPCKEATVSMETGDLCAOTAPTPKSEL	782
Dd	552	GMPGHTTHMIEHIGSLGLYHFRISILIQSCDPCKMETBSPETGDLCDNTPNAPKHS	611
QY	783	CREPEPISDTGCFTRFPGAPFTNYMSYTDNCTDFTFNQVARNHCYLDLYYQOWTSRK	842
Dd	612	CGDRPGNDTGFSHFNTPYNNFMSYADDDCTOSFTFNQVARNHCYLDLYYQOWQPSRK	671
QY	843	FTPIPIPMVYIGQTNKSLTIHMLPISGVVYDRASGSLCAGCTEGTGTROYVHTSSSRV	902
Dd	672	PAPVALAQVGLHTTSDTYLEMPFPIIDHFERELEGSCHLCSBERILVQVANSASSMP	731
QY	903	CDSSGYTPPEBAVCPDNDQPCBSLQAMSPREVLVHMMNTVPCP--TEGSLLELFOHPV	961
Dd	732	CSPSGHMSPRABEGHPDVEQCKSSVTRTMSPSAVNPHTVPBACPCEPOGYALELEFYPL	791
QY	962	QADTLTLWVT--SPFMESSQVLPFDEILLKMKESHGLRPLDFCQDILITIKL--HYDGVS	1018
Dd	792	VPESTLITWTVSTJDMWSSGAVNDIKLAVSGKNSLDEPQNVFCVPLITIRLMDVGEERY	851
QY	1019	GVKYVTPFERIEIDAALLTQSHPSPSCGCRPRVRYQVLRDPPFASGLRVVVVTHSHRKETD	1078
Dd	852	GIQIYTLDEHLEIDAMLTSTADPTLCQCKRYKXVVRDPLQMDVASIL--HLMRKVD	910
QY	1079	VEVTPRGQYQYQVLAAGGELGASPRPLNHTHGAPECGDGKYSERLGECCDGLVSGDG	1138
Dd	911	MDLIMGSYAYQWVYIISGTESEBPBAVYTIHGRCYCBDDGIIQXQOGEQCDMMKINDG	970
QY	1139	CSKVCELEEGFNVCVEPSLCYWYBEGDGCEPERKTSIVDGIYTPKCYLDOUMATRAVSS	1198
Dd	971	CSLFRQSVSRNCIDBPSRCYFHDDQCEBEQKTSIKDCGYVTPQGFLODMASNAVS	1030
QY	1199	HEDKKKCEVSLVTGER--HSLICTSYHPDLPNHRPLTGMFPCVASENETQDDRSQPECSL	1257
Dd	1031	HQD--QCCGQWYIIGQPAASQVCKTKVIDLSEGISQHWMPCTISYPSQ-----	1078
QY	1258	KKEDEVMLKVCENRGEARAFILPTTQGVLVGEHQPTVILYLTDVAGSNHSLGTGSL	1317
Dd	1079	LAQTTFMRAFSQWVAAAVVHVLVTGTGYGDKQKETSISQVLLDPTQDSHIDGLHVL	1138

[illegible]



A/Molecule type: mRNA  
 A/Residues: 26-584 <HOU>  
 A/Cross-references: EMBL:X1362; NID:930197; PIDN:CAA32541.1; PID:G736240  
 A/Experimental source: clone CRI-4  
 R/Klickstein, L.B.; Barlow, T.J.; Miletic, V.; Rabson, L.D.; Smith, J.A.; Fearon, D.T.  
 J. Exp. Med. 168, 1699-1717, 1988  
 A/Title: Identification of distinct C3b and C4b recognition sites in the human C3b/C4b  
 A/Reference number: S03843; MUID:89035992; PMID:2972794  
 A/Accession: S03843  
 A/Status: translation not shown  
 A/Molecule type: mRNA  
 A/Residues: 1-894, 'A', 896-1000, 1451-2064, 'T', 2066-2276, 'P', 2278-2299, 'H', 2301-2325, 'T', 2327-2  
 A/Cross-references: EMBL:Y00816; NID:930185; PIDN:CAA68755.1; PID:G30186  
 R/Klickstein, L.B.; Wong, W.W.; Smith, J.A.; Weis, J.H.; Fearon, D.T.  
 J. Exp. Med. 165, 1095-1112, 1987  
 A/Title: Human C3b/C4b receptor (CR1). Demonstration of long homologous repeating domain  
 A/Reference number: A28507; MUID:87168191; PMID:2951479  
 A/Accession: A28507  
 A/Molecule type: mRNA  
 A/Residues: 953-1221, 'FV', 1224-2064, 'I', 2066-2276, 'P', 2278-2299, 'H', 2301-2325, 'T', 2327-2  
 A/Cross-references: GB:X05309; NID:930196; PIDN:CAA28933.1; PID:9809019  
 R/Wong, W.W.; Klickstein, L.B.; Smith, J.A.; Weis, J.H.; Fearon, D.T.  
 Proc. Natl. Acad. Sci. U.S.A. 82, 7711-7715, 1985  
 A/Title: Identification of a partial cDNA clone for the human receptor for complement fr  
 A/Reference number: A94073; MUID:86067975; PMID:2933745  
 A/Accession: A24748  
 A/Molecule type: mRNA  
 A/Residues: 311-333; 729-745; 831-845 <MO2>  
 A/Cross-references: GB:M11569; NID:9180991; PIDN:AAA52297.1; PID:9180995; GB:M11617; NID  
 C/Genetics:  
 A/Gene: GDB:CR1; CD35  
 A/Cross-references: GDB:119800; OMIM:120620  
 A/Map position: 1Q32-1Q32  
 A/Intron: 41/1; 101/1; 134/2; 163/1; 296/1; 355/1; 389/2; 418/1; 491/1; 551/1; 584/2; 6  
 ; 184/2; 1513/1; 1646/1; 1705/1; 1733/2; 1768/1; 1844/1; 1904/1; 1937/2; 1966/1; 2099/1  
 C/Superfamily: C4b-binding protein alpha chain; complement factor H repeat homology  
 C/Keywords: duplication; glycoprotein; transmembrane protein  
 F/1-41/Domain: signal sequence #status predicted <SIG>  
 F/42-2489/Product: complement C3b/C4b receptor, membrane-bound form #status predicted <K  
 F/42-584/Product: complement C3b/C4b receptor, secreted form #status predicted <MATS>  
 F/43-99/Domain: complement factor H repeat homology <FH01>  
 F/144-161/Domain: complement factor H repeat homology <FH02>  
 F/166-232/Domain: complement factor H repeat homology <FH03>  
 F/238-293/Domain: complement factor H repeat homology <FH04>  
 F/297-353/Domain: complement factor H repeat homology <FH05>  
 F/358-416/Domain: complement factor H repeat homology <FH06>  
 F/421-487/Domain: complement factor H repeat homology <FH07>  
 F/493-549/Domain: complement factor H repeat homology <FH08>  
 F/554-611/Domain: complement factor H repeat homology <FH09>  
 F/616-682/Domain: complement factor H repeat homology <FH10>  
 F/688-743/Domain: complement factor H repeat homology <FH11>  
 F/747-803/Domain: complement factor H repeat homology <FH12>  
 F/808-866/Domain: complement factor H repeat homology <FH13>  
 F/871-937/Domain: complement factor H repeat homology <FH14>  
 F/943-999/Domain: complement factor H repeat homology <FH15>  
 F/1004-1061/Domain: complement factor H repeat homology <FH16>  
 F/1066-1132/Domain: complement factor H repeat homology <FH17>  
 F/1138-1193/Domain: complement factor H repeat homology <FH18>  
 F/1197-1253/Domain: complement factor H repeat homology <FH19>  
 F/1258-1316/Domain: complement factor H repeat homology <FH20>  
 F/1321-1387/Domain: complement factor H repeat homology <FH21>  
 F/1393-1449/Domain: complement factor H repeat homology <FH22>  
 F/1454-1511/Domain: complement factor H repeat homology <FH23>  
 F/1516-1582/Domain: complement factor H repeat homology <FH24>  
 F/1588-1643/Domain: complement factor H repeat homology <FH25>  
 F/1647-1703/Domain: complement factor H repeat homology <FH26>  
 F/1708-1766/Domain: complement factor H repeat homology <FH27>  
 F/1771-1837/Domain: complement factor H repeat homology <FH28>  
 F/1846-1902/Domain: complement factor H repeat homology <FH29>  
 F/1907-1964/Domain: complement factor H repeat homology <FH30>  
 F/1969-2033/Domain: complement factor H repeat homology <FH31>  
 F/2041-2096/Domain: complement factor H repeat homology <FH32>  
 F/2100-2156/Domain: complement factor H repeat homology <FH33>

F/2161-2219/Domain: complement factor H repeat homology <FH34>  
 F/2224-2290/Domain: complement factor H repeat homology <FH35>  
 F/2298-2354/Domain: complement factor H repeat homology <FH36>  
 F/2359-2415/Domain: complement factor H repeat homology <FH37>  
 Query Match 3.0%; Score 292.5; DB 2; Length 2489;  
 Best Local Similarity 19.1%; Pred. No. 3.5e-10;  
 Matches 390; Conservative 211; Mismatches 691; Indels 745; Gaps 109;  
 76 GNTLRPVVEGEIHHNRGSKPDTEGNAVLVPRDLLENAGLRGAEEPAAPVGGSP 135  
 434 GKPLEVFPFG-KAVNYTCDBHPD-RGTSFDLI-----GESTLR-CTSDQNGCWSSPA 484  
 136 GQSEILD--DDAYLGNQSKESLGRAGIOKGSAMAATTTA----IFTLLN-EKRPET 187  
 485 PRGCIILHCAAPDFL-----FAKTKQTNSDPFGISLKECEPXY 527  
 188 QRRGMAKSRQRYVKKRAEDGGGDSGISHPQWPGRSLKHRYKSPPEESN----- 240  
 528 YGRPFSTICDNLVWSS-----PDVCKRKSKTPPDPPVNGMVHVT 569  
 241 --QNGEGSIREAETPSQVG--LPITYSGRERLIREVLAEIREAFTVAMVKPE 296  
 570 DIVGSRINY-SCITGHRLLGHSSABCLISGNAMWTKPPICORIPCGE----- 618  
 297 GGQNNPAIINGVDFNCSHTVSDKGMALGIRGDKGRDARFPFSLCTDRYKATILISH 356  
 619 ---PPTIANGDFTSTIRENFHTGVSVTYCNQSGGRKY---FEL---VGEPISTCTS 667  
 357 SRVQPTGTHVA-----ATYDGRMALVYDGTQVASSLDQ-----SGP-- 394  
 668 NDDQVGWSPAPQCIIPNKCTPPNVENGILVSDNRSLFSLNEVEFRCPGPFVWKGPBR 727  
 395 ---LN--SPFMASCSL-----LIGGD-----SSEGGHYFRHLG 424  
 728 VKCQALMKWPEPEPSRCRVQCPDPDVLHARTQDKDNFPGQEVFSCGXYLRGAAS 787  
 425 TLVF---WSTALPQSHFGHSSQSSGSEBATDVLVLAEPVNTW---VPRDEKYPR 477  
 788 MRCPTGQDMSPAPD-----TCEVKSDDPWGQLNGRVLPVNLQAKVDFVCDGDFQ 841  
 478 LE-----VLQGE-----PEPEIL-----SPLOPLCGQTV---CD 505  
 842 LKGSASVCLAGHESLMNSVPCBOIPCPSPVIVNGRHGKPLVFPFGKTVNTCD 901  
 506 -----NVELIS-----QYNGWPLRGEKVIRYOVNVC--DEGLNPIVSEBQ 546  
 902 PHPDRGTFDLIGESTIRCTSDPQNGW---SSPAPRCGILGHQAPDHLFAKLTQT 958  
 547 -----ILQHEALNEAFSR-----YNISQQL-----SVHGV 572  
 959 NASDFPIGTSLKTECRPEYVGRPESTICLDNLVWSSPKRVCKRKSCTPPDPVNGMVHVI 1018  
 1019 TDIGVGRIT-----NYSC--TTGHRLLGHSSABCLISGNAMWTKPPICORIP 1065  
 632 CNNMLNDFDGDCCDPQVADVKTCTPDPSPKRAYSVKELKEALQNSTHFLNIYFAS 691  
 1066 CG-----LPPIA-----NGDPIST 1080  
 692 VREDLAGAATPMDKXAVTHLGGIVLSPAYYGMPTGHTDTHIEVGHVILGLY-----H 743  
 1081 NRENF-----HYGSVY---TRCNNGSGGRKVFELVGEPSITCTNSDDQVG 1123  
 744 VFKVSERESCNPKCTEVSMTGDCADTAFT-----PKSELGR 784  
 1124 IMSGPAQ--CIIPNKCTPPNVENGILVSDNRSLFSLNEVEFRCPGPFVWKGPBRXKCQ 1181  
 785 ---EPEFTSPTCGTFRPGAPFTNMYNSYTDNCTDNFTPNQVAMHCYLLDLVYQQTESR 841  
 1182 ALNKEPELPSCSRVCVP--PDVLHAERTQD--KDNPSFGQEVFVSC----- 1226

```

Qy 842 KPTPIPIPVNIGQTNKSLTIHMLPPIGSLVVD-RASGSLGACTEDGTFRQYVHTASSR 900
Db 1227 BPG-----YDKRGAASM--RCTPOGDWSPAPPTCEVK 1256
Qy 901 RVCSSSGWTPBEAVGPPD-----VDOPCEPSIQAMSPVHLVHMNTVPCPTBGSGL 954
Db 1257 SCDFPMQQLNGRVLFPVNLQAKVDPVCEBGR-----LKGSA- 1297
Qy 955 LLFQHPQADTLTLMVTSFFMSSOVLPTEILENKSVHLG-PLDTF-----CD 1004
Db 1298 ---SYCLAGMESLNNSSVPV-CEQIFCPSPVLPNGR--HTGKPLEVFPFGKAVNTCD 1351
Qy 1005 IPLTIKLHVDCVSGVAVTPEDERIEIDALLTSOPH-----SPL-----CSGCRP 1050
Db 1352 ---PHPRGTS-----FD-LIGESTRICTSDPGNGVWSPAPRCGILGHQAADH 1398
Qy 1051 VRYOVLDRPPASGLPVVVTHTSHRKFTVEVTPGOMVOYOLAERG-----ELGE 1101
Db 1399 FLFAKLTKQTNASDFPI-----GTSLEYECRPEYGRPFSTICLDNLVW 1442
Qy 1102 ASP-----PLNHHGAPY-CGDGKVSERLGECDGDLVSGDGSKVCELEBGF 1149
Db 1443 SSPADVCCKRCKTPRDPVNGMHTITDIOGSRINSCCTGHRILGHSSAB----- 1494
Qy 1150 NCVEPRLCYMYEEDGICEPPEKRTSIVDCG-----IYTPKGYLDQMATRAYSSH 1199
Db 1495 -CISGNTAHMSTKPIQCR-----IPGILPTIANGDPISTNREHFHGVVYTRCN 1546
Qy 1200 EDKCKCPVSLVTGERPHSLICTSYHND-----LPMH-RPLTWCFCVASENETQ 1246
Db 1547 LGSGRKFKFELVGBR-SIYCTSNDOVGWISGAPQCIIPKCTP-----PNVERGILVS 1600
Qy 1247 DRESEOPRSLKKEDEVVLKTCFNRPRGARAFIFLTITDGLVPGHQOFTVLTLTVRG 1306
Db 1601 DNRS---LFSLNEVVERFCQPGFWKGRK---VKQALNKEPELPPSCSRV----- 1646
Qy 1307 SNHSLGTGYSLCOHNPILINVT---HQNVLPHHTSVLNFSSBRVGISAVALTSSRI 1363
Db 1447 -----CQPPPEILHGEHTRSHQD-----NFSR----- 1668
Qy 1364 GLSAPSNISDEGONHOGQSCIRPCKQ-----DSCPSLL-LDHADVNC 1409
Db 1669 QGEVFTSC---EPGYDLRGAASLH--CTPOGDWSPBAPRCVKSDDDELQGLPHGRVLP 1723
Qy 1410 TSIPEGMLKCAITOCRFALOGSGQYRPMQKEILLCCSSGHDQNVN--CLPVDGCPD 1468
Db 1724 LNLQLG-AKVSFVCDGFRLLKSSVSH-----CVLVGKRSLLMNNVSPVCEHIFCENP- 1774
Qy 1469 PSLVNYANFSCSEG-TKFLKRCSSISCV-----PAKLOGLSPMLTCLD-----GLMSLP 1517
Db 1775 PALINGRHTGTPSGDIPIYKGEISYTCBHPDKMTFNLIGEST-IRCTSDPHNGVWSSP 1833
Qy 1518 EYVYCKL-----ECDAPIILANLILPHCLQD-NHDVETICKYCKPGYVVAESAEGKVR 1571
Db 1834 APRCELTVRACHCTPEQPPASPTIP--INDFEFVGTSLNTECRPGYF----- 1881
Qy 1572 NKLKTIQCLEGGIWE--OGSCIPVYCEBPVPVFEEM-----YECTNGFSL- 1614
Db 1882 GKMSISICLNLVMSVSDNCRKSCGPRPPEFGMHNINIDTQGSVWVYNSCNEGFALI 1941
Qy 1615 ---DSQCLANCOBERKLPILCTKEGLWTQBFKLCENLQGBRPPSELNS----- 1662
Db 1942 GSPFTTCLVSGNNV-----TWDKKAPICEII--SCBPPPTISNDPFSNNRTS 1987
Qy 1663 ---VEYKCEQYG-----IGAVCSP-----LCYTP-S 1686
Db 1988 FHNGETVYVYQCHTGDGQOLFELVGBERSIYCTSKDOVGWVSSPPRCISTNKCTAPBE 2047
Qy 1687 DPMVLPEVIT---ADTLEHMMBP-----YKQSVICTGRQWHPRPVVLHGIQSGEP 1735
Db 2048 NAIRVPGNRSPFSLTEIVRFRQCPGFVWVGSHTVQCCQNGRM--GPKLPHCSRVCP 2102

```

```

RESULT 3
136936
C:Species: Pan troglodytes (chimpanzee)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 21-Jul-2000
C:Accession: 136936; 136937
J:Birmingham, D.J.; Shen, X.P.; Hourcade, D.; Nickells, M.W.; Atkinson, J.P.
J:Immunol. 153, 691-700, 1994
A:Title: Primary sequence of an alternatively spliced form of Cr1. Candidate for the 75,.-
A:Reference number: 136935; MUID:94292799; PMID:8021505
A:Accession: 136936
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-2014 <BIR1>
A:Cross-references: GB:124920; NID:g551564; PIDN:AAA51438.1; PID:g557725
A:Accession: 136937
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-397,1751-2014 <BIR2>
A:Cross-references: GB:124921; NID:g557726; PIDN:AAA51439.1; PID:g557727
A:Gene: Cr1
A:Superfamily: C4b-binding protein alpha chain; complement factor H repeat homology
C:Keywords: duplication; glycoprotein; tandem repeat
F:18-74/Domain: complement factor H repeat homology <FH01>
F:79-136/Domain: complement factor H repeat homology <FH02>
F:141-207/Domain: complement factor H repeat homology <FH03>
F:213-268/Domain: complement factor H repeat homology <FH04>
F:272-328/Domain: complement factor H repeat homology <FH05>
F:336-391/Domain: complement factor H repeat homology <FH06>
F:396-462/Domain: complement factor H repeat homology <FH07>
F:470-526/Domain: complement factor H repeat homology <FH08>
F:529-586/Domain: complement factor H repeat homology <FH09>
F:722-778/Domain: complement factor H repeat homology <FH10>
F:1041-1107/Domain: complement factor H repeat homology <FH11>
F:1172-1228/Domain: complement factor H repeat homology <FH12>
F:1233-1291/Domain: complement factor H repeat homology <FH13>
F:1296-1363/Domain: complement factor H repeat homology <FH14>
F:1432-1489/Domain: complement factor H repeat homology <FH15>
F:1625-1681/Domain: complement factor H repeat homology <FH16>
F:1749-1815/Domain: complement factor H repeat homology <FH17>
F:1823-1879/Domain: complement factor H repeat homology <FH18>
F:1884-1940/Domain: complement factor H repeat homology <FH19>
Query Match 2.98; Score 289; DB 2; Length 2014;
Best local similarity 18.24; Pred. No. 4.4e-10;
Matches 317; Conservative 173; Mismatches 498; Indels 754; Gaps 92;
Qy 599 ECEHPLTGYDGDGR--LQGRYS-----WNRRDGLCHVCNNMLNDFDDGCCDPQ 648
Db 35 EFERPITVILNVECRPPGYRPSIILKNSVMTGAQRDR----- 75
Qy 649 VADVRKTCFDPDSKRAYMSV-KELKEALQI--NSTHFLNITRASSVREDLAGATYPMW 705
Db 76 ---RKSCRNPDPVNGMVMHTIDQPGSQIKYSCITGYALISSSTATCIISG-DYIWD 130
Qy 706 KD-----AVT-----HLGGLVSPAYYGMFGTDMIHVEGVHLG 740
Db 131 NETPTICRIPOGLPPTTNGDFISTNRENHYGSV---TYRCNPGSGGKRVFELVGEPS 187
Qy 741 LY-----HVFKVSERSCNDPCKETVSMETGLCADT----- 774
Db 188 IYCTSNDOVGWISGAPQ--CIIPNKCTPPNVNENGLVSDNLSFLNEVEVERCQPGF 245
Qy 775 ---APTPK-----SELCEP----- 786
Db 246 VMKGPVVKCOALNKEPELPSGRVCOPPDVLAERTORDKDNFSPGQEVFYSCEPGY 305
Qy 787 -----EPSTDTC-----GTPRFP-----GAPFTYMSYTD 811
Db 306 DLKGAASLRCTPGDWSPPRPTCEVKSDDPMGQLNGRVLFPVNLQLAGK-----VD 358
Qy 812 DNCTDNFTPNQVAMHCYLDLVVQOWTESRKPPI-----PIPVNIGQTNKSLTIHML 865

```

```

Db      359 FVCEGQKLGSSASVCLAGMESLWSS---VPVCEQIFCPSPIVPLNGR-----HTG 409
Qy      866 PRISGVVYDASAGSLGCACTEDGTFRQYVHTASRRVCD--SSGYW-TPREAVG----- 916
Db      410 KPLEVFFPGAAVNTCPHPDRGTTFDLIGESTIRCTSDPOGQVWSSPAPRCIILHCQ 469
Qy      917 PPD-----VDQCEPSLQAMSPREVLVHMMVTPC-----PTGCSLEL 955
Db      470 APBHFPLAKLKTQNNASDPFICTSLK-YKCRPEYGGPFSITCLDNLWSSPKVCKRKS 528
Qy      956 LFGHPVADTLTLMTVTFMES----- 977
Db      529 CKTPPDVNGKWHVITDIQVGSRIWVSCTGHRILGHSSAECILSNAAMWSTKPICOR 588
Qy      978 -----SQVLPDTLELNKKSVHLGPLDTF----- 1002
Db      589 IPCGLPPTIANGDITSTNRENFHGSVYVTKCNPGSGRKVFELVGEPSIYTSNDQVG 648
Qy      1003 -----CDIPLTIKLHVDGKVSQVY-----YTFDERLEIDA----- 1033
Db      649 IWSGPAPOCIIF-NKCTPRVVENGIIVSDNRLFSLNEVVERFCQPGFAMKGRVYKQ 706
Qy      1034 ALLTSQPHSLCSG-CRPVRYQVLRDPPFASGLFVVVTHSHRKTVE-VTPGQMYQYQV 1091
Db      707 ALNMKEBELPSCSRVCOF-----PPDVLHARTQDRKDNFSPQGEVYFS- 750
Qy      1092 LAEAGSBL-GEAS-----PRLNHIGAPYCG-----DG-----KYSERLGE- 1127
Db      751 -CEGVDYLRGAASIRCTPQGWSPAAPTCVEKSCDGMGOLLNGRLVFPVNLQIGAKVDF 809
Qy      1128 -CDGDGLVSGDGS-----KVCEL----- 1145
Db      810 VCBGDFOLKSSASVCLAGMESLWSSVPCBDFPCSPVITNGHTGKLEVFPRGK 869
Qy      1146 -----EEGFVCEBPSLCYMEBGDI-----CE----- 1168
Db      870 AVNYTCDPHPRDRTTFDLIGESTIRCTSDP-----QNGVWSSPAPRCGILGHQAPDH 923
Qy      1169 -----PFRKTSIVDCGIY-----PKGY-----LDQMAIRAASSHED---KKKC- 1205
Db      924 FLFAKLKTQTNASDPFICTSLKRYCRPEYGGPFSITCLD---NLWSSPKDVCRRKSC 980
Qy      1206 -PVSILVGEPSHLI-----CTSYHPLDPNHR-----PLGMWF 1236
Db      981 TPRPVVGMWHVITDIQVGSRIWVSCTTGH-RLIGHSSAECILSNAAMWSTKPICOR 1039
Qy      1237 PC-----VASENETQDPRSSEQPESGLKKEDEVMLKVCFNRPGEARALF-----IFLTT 1284
Db      1040 PCGLPPTIANGDITSTNRENFHGSV-----VTYCNLGSRGKRVFELVGEPSIYCTS 1092
Qy      1285 DG-----LVPEHQOQPTV--LYLTIVRSNHL----- 1311
Db      1093 NDQGVIGWSGAPACIIPNCKTPPVENGLIVSD---NRLFSLNEVVERFCQPGFVMK 1148
Qy      1312 GTVGLSCQ-----NHPLIINTVHQNVLFH--HTTSVLLNFSSPRVIGSAVALRTSSR 1362
Db      1149 GPHVVKQALMKWPELPSGRCVCOPEIILHGETSHQDNFSP----- 1193
Qy      1363 IGLSAPNSCISEDEGQNHQSGCIHRPGQK-----DSCPSLL--LDHADVYN 1408
Db      1194 -GGGVFVYSC--BFGYDLRGAAHLH--CTPGQDWSPEAPRCVASCDFGLQPLHGRVLF 1247
Qy      1409 CTSGIPBLMKALITCQGFALQASSGOYIRPMQKEILLTSSGHWQDNVS-CLFVDCGVP 1467
Db      1248 PLNLQQL-ARVSPVCEGFRLLKSSVSH-----CYLVGMRSLMNNSVPCBDFICFNP 1299
Qy      1468 DPSLVYANFSCSEGTFF-----LKRCSISCP-----PAKIQGLSPMLTCLD-----G 1512
Db      1300 -PALINGRH-----TGTFRGDIPLYGKEISYTCDPHRDKMTNLLGEBF-ICTCTDPRHNG 1353
Qy      1513 LMSLPEVYCKL-----ECDAPIILNANLLPHCLQD-NHDVGTICKYCKECPGYVAESA 1566

```

```

Db      1354 WSSPAPRCLEPVNAGHCKTPEQFPFASPTIP--INDFEPPVGTSLNVECRPGYF----- 1406
Qy      1567 EGKVRNKLTKIOCLEGGIWE--OGSCIPVCEPPPPVFEQM-----YECTN 1610
Db      1407 -----GKMFISLLENLWSSVEDNCRKSCGPPPEPFNGMWHINTDQSGSTNYSCNE 1461
Qy      1611 GFSL-----DSQVLANCQNEREKLPILCTKBLWTOEFKLCENTOGCEPPPSSELSN---- 1662
Db      1462 GFRILGSPSTTCLVSGNNV-----TWDKAPICEII--SCPEPPTISNGDFYS 1507
Qy      1663 -----VEKCEQGYG-----IGAVSP-----LCV 1662
Db      1508 NMFSPHNGTIVTYOCHTGPDGOELVELVGERSIYCTSKDQVWSSPPPRCISTNKT 1567
Qy      1683 IPP-SDPVMLEPENIT-----ADTLEHMMEP-----VKQSIYCTGRORHPPVLVHCIOQC 1733
Db      1568 APEVENAIRVGNRSPFSLTEIVFRCPQGFVWGVSHTTVOCTQNGRW--GPKLPHCSRVC 1625
Qy      1734 EP 1735
Db      1626 QP 1627

RESULT 4
MEGF6 protein - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jul-2000
C:Accession: T13954
R:Nakayama, M.; Nakajima, D.; Nagase, T.; Nomura, N.; Seki, N.; Ohara, O.
Genomics 51, 27-34, 1998
A:Title: Identification of high-molecular-weight proteins with multiple EGF-like motifs
A:Reference number: 214126; MUID:98360089; PMID:9653030
A:Accession: T13954
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1574 <NA>
A:Cross-references: EMBL:AB011532; NID:G3449293; PIDN:BA32462.1; PID:G3449294
C:Genetic8:
A:Gene: MEGF6

Query Match      2.9%; Score 282.5; DB 2; Length 1574;
Best Local Similarity 19.7%; Pred. No. 8.1e-10;
Matches 297; Conservative 81; Mismatches 476; Indels 657; Gaps 86;

Qy      571 QVNSTLNRHV-----LVNCEPS-----KIGNDHCEDECHPLTYD 608
Db      169 RAHNGGCGHRCVNTPGSYLCECKRPFRLHTDGRCLAISGCTLGNGGCGHQVQLTVTH 228
Qy      609 GGCRLQGRCYSMNR-----DGLCHVECNMNLNPDGDDCCD--QVADVRYKTC 656
Db      229 RCQCRPQYQLEDDGRRCVRRSPCAEAGGCGMHIQBELRGLAHCG--CHPEYQLAADRKTC 286
Qy      657 FDDPSKRAYSVYKELKALQNLSTPLNLYFASVVEDLAGAATMPWMDAATVHLGIV 716
Db      287 EDVDE-----CALGLAQCHAGCLNTGSKFYCH----- 315
Qy      717 LSPAYYGMGHDTMTIHVGHVIG--LVYH-FKGVSRSESCNDCKETVBSMETGDL 770
Db      316 -----AGYELGADRGOCYRIEMELVNSCEAGNGGCSHGCSHTSGPL 357
Qy      771 CA-----DTAPFKSELCREPEPSDTCGTRPPG-----APFTMYMYTD 811
Db      358 CTCPRGYELDEDKTCIDIDCANSPCCQ-----ACANT--PGGYECSFAGYRLNTD 409
Qy      812 -----DNCT-----DNFTPNQVAMHCYLDLVYQMTESRKPPI----- 846
Db      410 GCGCEDVDEGASGSGGGEHHCNSLNASGFCGCEAGYVLDDBRGCTLESVVDLQRLP 469
Qy      847 ---PIPMVIGQTNKSLITLHLPPI-----SGVYDRASSGLCGACTEDGTFRQYVHTASS 899
Db      470 FVRPLPIHIVLRDE-----LPRLFDQDYGAEEAAMAEIRGE-----HTLVE 511

```

```

Oy 900 RRV-----DSSGYTPBEAVG-----PPV-----DQP 923
Db 512 KPVLDHDSFGHDSITCDCCANGGTCFPGDGCDCPCBGMTIICNCTCPPTPGKNGSSP 571
Oy 924 -----CEPSLOA--WSPVH----- 936
Db 572 CTCNGGTCDDVLAACRCPGVSGAHCEDEGCPKFGYKHKRCKKCANRGRCHRYGACL 631
Oy 937 -----LYHNMNTVPCPT-----EGCSLELLFQHPVADTLTMTWTFEMSSQVLDTEILL 988
Db 632 CDPELYGFCHLACPMWAFGPGCS-----EDLTC 660
Oy 989 ENKESVHLGPDFTCDIPLITKLHVDGKVSQVYTFDERIEIDALLTSOPHSPLCSGC 1048
Db 661 EGSHTRSNPKDGC-----SCKAGQG-----ERQABESGFPF-----GC 699
Oy 1049 RPRVRYQVLRDPPFASGLRPVVVTHSHRKPTDVEVTPGMYQYQVLAAGELGEASPPILNH 1108
Db 700 R-----CTCQPG-----VACDPVSGECRTQCP----- 724
Oy 1109 IHGAPYCGDGKVSRLGECDDGLVSGDGSKVCELEGNCVGEPSLCTMTEGDGICE 1168
Db 725 -----GYQGEDCCGCEPVGTF--GVNCSGSC-----SCVGAAP--CHRYTGECLCP 765
Oy 1169 PFERKTSIVDCGIYTPKGYLDQMATRAVSSHDKKKCPVSLVTEGPHSLICTSYHPLDPN 1228
Db 766 P--GKTG--EDCGADCPBG--RWGLGC-----QETCAC-----EHASC----- 799
Oy 1229 HRPLTWMPFVCASENETODDSEDEPSGLKKEDEVMLKVCNRRPGEARAFIFLTDTGLV 1288
Db 800 -NPEFG-----TCLCPG-----FV----- 813
Oy 1289 PGEHQPTVTLVLTVDVRSNHSLSCTYGLSCQHNPLIIVTHQNVLFHHTSVLNFSSP 1348
Db 814 -GSRQDTC-----SAGWYGTGCGQIRCAANDGH-----CDP 844
Oy 1349 RVGISAVALTSTSRIGLSAPNSCISEDQNHQSGCITHPC-----GXODSCPSLLD 1402
Db 845 TTGRSCA--PGMTGLSCQRAAC--DSG--HWGDDCH--PCNCSAGHNCDAVSGCLIC 895
Oy 1403 HADV-----VNCTS--IGPGL--MKCAITCORGFALQASSGOYIRPMOKELLTSSGHW 1453
Db 896 EAGYGPCEQSCRGYGYPSCEQKC--RCEHGAACDHVSG-----ACTCPAG--W 942
Oy 1454 -----DQVNSTLPVDCGVPRPSLVNANFSSSEGTKTLKXCSISCVPR 1497
Db 943 RGSFCEHACPAFGFGLDCDSAC--NCSAGAPCDAVTGSICCPAG--RWGPRADGSCPLT 998
Oy 1498 KLGSLSPWLTCLB-----DGL-----WSLPEVYCKLECDAPPIILNANL----- 1536
Db 999 FGLNCSQJCTGPNAGSCDSVYTGQCHCARGMGP--CLQAC--PGLYGRKNCQHSCLCRN 1054
Oy 1537 -----LLPHCLQDNHDVGTICKYECKRGYVVAASABGRVKNKLKIQCLEBGI----- 1584
Db 1055 GGRCDPILGQCTCPBGMTGLACENECILPQHVAAGQ-----LNCCLHGIGICDLRLTG 1106
Oy 1585 -----WE-----QSSCIP-----VVCERP-----PVEBGM--EC 1608
Db 1107 HCLCPAGWTGDKQSSCVSGTFGVNACEBHACRKGASCHNYVACFCPPGMWGPCEQAC 1166
Oy 1609 TNGFSLDQCVLNCQBEREKLPICTKKGMLTQBERKLCENLOGEPRPSELN--SVEKVC 1667
Db 1167 PRGM--FGBAACARC-----LEPTNA-----SCHNVTGBCRCPPGFTGLSCQAC 1209
Oy 1668 EOGYGAIGAVCSPLCVIPSPDPVMLPENITADTLEHM--MEPVVQSVICTGRROMHPDVL 1726
Db 1210 QPG--TFGDCHLCCQCPG-----ETWACDPA--SGVCTCAAGHYGTGL 1250
Oy 1727 VHCIO-----SCEPF-----QADWCCTINNRAYC--HYDGGCCSSTLSSKVIYPPAALCD 1776
Db 1251 QRCPSGRYVPGCEHICKLNGGTCDPATGACVCPAGFLGADSLACPOGR--FGPSC- 1305

```

```

Oy 1777 LBECTCRDPRKA 1787
Db 1306 AHVACACRQGAR 1316

RESULT 5
A:30359
P-selectin precursor - human
N:Alternate names: CD62 antigen; granule membrane protein 140
C:Species: Homo sapiens (hmn)
C>Date: 18-Oct-1989 #sequence_revision 30-Sep-1991 #text_change 16-Aug-2002
C:Accession: A30359
R:Johnston, G.I.; Cook, R.G.; McEver, R.P.
Cell 56, 1033-1044, 1989
A:Title: Cloning of GMP-140, a granule membrane protein of platelets and endothelium; see
A:Reference number: A30359; MUID:89168432; PMID:2466574
A:Accession: A30359
A:Molecule type: mRNA
A:Residues: 1-830 <JOH>
A:Cross-references: GB:M25322
A:Note: parts of this sequence, including the amino end of the mature protein, were confi;
C:GeneticB:
A:Gene: GDB:SELP, GRMP
A:Cross-references: GDB:120018; OMIM:173610
A:Map position: 1q22-1q25
C:Superfamily: P-selectin; C-type lectin homology; complement factor H repeat homology; i
C:Keywords: cell adhesion; glycoprotein; phosphatidyl; phosphoprotein; surface antigen
F:1-41/Domain: signal sequence #status predicted <SIG>
F:142-830/Product: P-selectin #status experimental <MAT>
F:163-194/Domain: EGF homology <EGF>
F:200-257/Domain: complement factor H repeat homology <FH01>
F:324-319/Domain: complement factor H repeat homology <FH02>
F:324-319/Domain: complement factor H repeat homology <FH03>
F:386-443/Domain: complement factor H repeat homology <FH04>
F:448-505/Domain: complement factor H repeat homology <FH05>
F:510-567/Domain: complement factor H repeat homology <FH06>
F:572-629/Domain: complement factor H repeat homology <FH07>
F:642-699/Domain: complement factor H repeat homology <FH08>
F:704-761/Domain: complement factor H repeat homology <FH09>
F:772-795/Domain: transmembrane #status predicted <TM>
F:796-830/Domain: intracellular #status predicted <CYT>
F:54,98,180,212,219,411,460,518,665,716,723,741/Binding site: carbohydrate (asn) (covale

Query Match 2.5%; Score 249; DB 2; Length 830;
Best Local Similarity 22.1%; Pred. No. 4.9e-08;
Matches 102; Conservative 59; Mismatches 160; Indels 140; Gaps 25;

Oy 1382 GQSCIH--RPGCKODSCPSLLDHDVNVNCTSIGGLMKCAITCORGFALQASSGOYIRPM 1440
Db 191 GRECEYVRECGEELPQHVLNCSHPLGNFSFN--SQCSFHTDGYQVNG-----PS 240
Oy 1441 QKELLTSSGHW--DQNVSLCPVDCGVPRPSLVNANFSSSEGTKFLKX--CSISC-- 1493
Db 241 KLECL--ASGITWNRKPPQCLAAC--PPLKIPERGMITLHSAKARFOHSSCSFSEBEG 295
Oy 1494 ---VPKLTQGLSPWLTCLBGLWSLPEVYCK--LEC--DAP-----PILNANL 1537
Db 296 FALVGPVEVQ-----CTASGVNTARAPYCKAVQCHLAPRSSTMDCHNPLTFA-- 345
Oy 1538 LPHCLQDNHDVGTICKYECKRGYVVAASABGRVKNKLKIQCLEGIMEQ--GSCIFVVC 1595
Db 346 -----YGSCKRECGQGY-----RVNG--LDMLRCTDSGHMSAPLPCEAISC 386
Oy 1596 EPPPVVEGMYECT--NGBSLDQVLCNQF--REKPIILCTKKGMLTQBERKLENT 1649
Db 387 EPLSPVHGSNDSPSRAFPQYDNCSPFCABEGFMAGADIWRCDNLGQWTAAPVCOAL 446
Oy 1650 QGECPPPSSELSVEYKCEQGY--IGAVCS----- 1679
Db 447 QCGDLPPV--NEARVNCNHPFGAFRYQSVCSFPCNGBLLVGSVLCATGMMNVP 503
Oy 1680 LCVIPSPDPVMLPENITADTLEHMPEPVKVS-----IVCTGRRO 1719

```

Db 504 ECAIPICTPLISLPQNGTMTG---VQPLGSSSSYKSTQCFICDEGYSLSGPERLDCRTSGR 559

Qy 1720 WHPDPVLVHCIOGCEPFOAD-GWCDTINNRR-----AYCHY 1753

Db 560 WTDSPWCBAIKCEPLFAPRQGSIDCDTGTGEFNVSTCHP 600

RESULT 6

E-selectin precursor - mouse

N/Alternate names: endothelial leukocyte adhesion molecule 1 (ELAM-1)

C/Species: Mus musculus (house mouse)

C/Date: 04-Mar-1993 #sequence\_revision 31-Dec-2000 #text\_change 02-Aug-2002

C/Accession: S23174; B42755

R/Becker-Andre, M.; van Huijsduijnen, R.H.; Losberger, C.; Whelan, J.; Delamarier, J.F.

Eur. J. Biochem. 206, 401-411, 1992

A/Title: Murine endothelial leukocyte-adhesion molecule 1 is a close structural and functional reference number: S23174; MUID:92283265; PMID:1375914

A/Accession: S23174

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-612 <BRC>

A/Cross-references: GB:M80778; NID:9193014; PIDN:AAA37547.1; PID:9193015

R/Weller, A.; Isenmann, S.; Westweller, D.

J. Biol. Chem. 267, 15176-15183, 1992

A/Title: Cloning of the mouse endothelial selectin. Expression of both E- and P-selectin

A/Reference number: A42755; MUID:92340574; PMID:1378846

A/Accession: B42755

A/Status: nucleic acid sequence not shown

A/Molecule type: mRNA

A/Residues: 'MKATAGV', 1-389, 391-612 <WEI>

A/Cross-references: GB:M87862; NID:9193107

A/Experimental source: endothelial cells

A/Note: sequence extracted from NCBI backbone (NCBI:P:109470)

A/Note: the sequence in Genbank entry M87862, release 117.0, (PIDN:AAA37577.1; PID:9193015) is incorrect. It is not selectin. The initiator is Met-1 or the AUG codon preceding that

C/Superfamily: P-selectin; C-type lectin homology; complement factor H repeat homology;

C/Keywords: glycoprotein; transmembrane protein

F/1-21/Domain: signal sequence #status predicted <SIG>

F/12-118/Domain: C-type lectin homology <LCH>

F/12-612/Product: P-selectin #status predicted <MAT>

F/143-174/Domain: EGF homology <EGF>

F/180-238/Domain: complement factor H repeat homology <FHI>

F/243-300/Domain: complement factor H repeat homology <FHI>

F/305-363/Domain: complement factor H repeat homology <FHI>

F/368-426/Domain: complement factor H repeat homology <FHI>

F/431-489/Domain: complement factor H repeat homology <FHI>

F/494-548/Domain: complement factor H repeat homology <FHI>

F/25,391,528/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 2.5%; Score 247; DB 2; Length 612;

Best Local Similarity 21.8%; Pred. No. 4.36-08;

Matches 121; Conservative 60; Mismatches 225; Indels 148; Gaps 27;

Qy 1233 TGMFPCVASENETQDRSEQEG-----SLKKEDE 1262

Db 20 TAMYVNASSELMYTDASACQRDYTHLVAIQNKEEINLYNSNKHSPSYWIGIRKYN 79

Qy 1263 VMLKVCNRPGEARAFIFLTLDLVPGE-----HQQPTVLYLTVDRGS----- 1307

Db 80 VMIWVGTKELTEBA-----QWAGGEPPNKRQNEDEIYIORTKDSGMNDECRNK 132

Qy 1308 -----NHSLGTYG-----LSQNHPLIINTHONVLF-----HRTTSLIN 1344

Db 133 KKLALCYTASCTVASCGBHGEIETINSYTCCKHGFGLGPNCEAVYCKQEHEDYSLN 192

Qy 1345 FSSPRVISAVALTSSRIGLS-----APSN-----CISEDGQHOGQSCIHRCGRKDS 1395

Db 193 CSHP---FGPSYNSSCSGFKGKRYLSSMETTYRCS-----SEMSAPACHAVE 242

Qy 1396 CPSLLLDHADVNVCTSIGPGL---MKCAITCGQFALQASSGQYIRPMOKKEIILLTSSG 1451

Db 243 CEALTHPAHGIRKCS--NPGSYPMNTTCTFPCVEGY-----BRVGAQNLCCTSSG 291

Qy 1452 HMD-QNVSCLPVDG-GVPDPSTLVNANFSCSEGT---KELAKCSISCVPAKLOGLSPW 1505

Db 292 IMNENPSCKAVYCDALPQD---QNGFVSCSHSTAGELAFKSSCNFTCEBSFTLQGAQ- 347

Qy 1506 LITLEDGLMSLPVEYCK-LECDAPPIILNANLLPHCLQDNHDVGTI CKYECKRGYVAE 1564

Db 348 VECSSAGQWTPQIPVCKAVQCEALSAPOGMMKCLPSASGPFQNGSSCFSCSEGFELKG 407

Qy 1565 SAEGYRANKLTKIQCEGGIWE---QGSCTFVCEPPPPVEGMYEC---TNGPSLSDSC 1618

Db 408 SR-----RLQCGPRGEMDSKKPTCSAVKCDVPRPQNGVWECHATTGETYKSSC 458

Qy 1619 VLNQNE---REKLPIICTEKGAMTQEFKLCENLQSECP---PPSELINSVEYKCEQGYGI 1673

Db 459 AFQNEBFLHSGAQLQECTSGKRTQEVPSQVQVQ---CPSLDVPGRKN---MCSGNAV 513

Qy 1674 GAVCSPLCVIPPSD 1687

Db 514 GTVCBFTC---PDD 524

RESULT 7

PL0009

complement C3d/Epstein-Barr virus receptor precursor - human

N/Alternate names: complement receptor 2; CR2/CD21

C/Species: Homo sapiens (man)

C/Date: 30-Jun-1992 #sequence\_revision 07-Jul-1995 #text\_change 22-Jun-1999

C/Accession: J10028; A39958; A32036; A24319; B24319; D24319; F24319; P101

R/Wells, J.U.; Tothaker, L.E.; Smith, J.A.; Wells, J.H.; Fearon, D.T.

J. Exp. Med. 167, 1047-1066, 1988

A/Title: Structure of the human B lymphocyte receptor for C3d and the Epstein-Barr virus

A/Reference number: J10028; MUID:8811282; PMID:2832506

A/Accession: J10028

A/Molecule type: mRNA

A/Residues: 1-1091 <WEI>

A/Note: nucleotides 1566-1625 are missing from Figure 1; therefore, residues 522-542 have

R/Moore, M.D.; Cooper, N.R.; Tack, B.F.; Nemerow, G.R.

Proc. Natl. Acad. Sci. U.S.A. 84, 9194-9198, 1987

A/Title: Molecular cloning of the cDNA encoding the Epstein-Barr virus/C3d receptor (com

A/Reference number: A39958; MUID:88097454; PMID:2827171

A/Accession: A39958

A/Molecule type: mRNA

A/Residues: 1-456, 'G', 457-644, 'R', 646-669, 'R', 671-816, 'NCSAEVILKWLIERAF', 835-840, 'L', 841

A/Cross-references: GB:J03655; NID:9181919; PIDN:AAA35784.1; PID:9181920

R/Fujiwara, A.; Hartley, J.B.; Frank, M.B.; Gruner, B.A.; Frazier, B.; Holgers, V.M.

J. Biol. Chem. 264, 2118-2125, 1989

A/Title: Genomic organization and polymorphisms of the human C3d/Epstein-Barr virus recep

A/Reference number: A32036; MUID:89123277; PMID:2563370

A/Accession: A32036

A/Molecule type: mRNA

A/Residues: 1-456, 'G', 457-658, 718-1050, 'I', 1052-1060, 'E', 1062-1091 <FUJ>

A/Cross-references: GB:J04463

R/Wells, J.U.; Fearon, D.T.; Klickstein, L.B.; Wong, W.W.; Richards, S.A.; De Bruyn Kops,

Proc. Natl. Acad. Sci. U.S.A. 83, 5639-5643, 1986

A/Title: Identification of a partial cDNA clone for the C3d/Epstein-Barr virus receptor

A/Reference number: A94114; MUID:86287311; PMID:3016712

A/Accession: A94114

A/Molecule type: protein

A/Residues: 226-330, 'XIQ', 257-267, 332-341, 583-591, 'Q', 593, 'D', 595-596, 728-735 <WE2>

A/Experimental source: B-lymphoblastoid cell lines SB and Raji

C/Genetics:

A/Gene: GDB:CR2

A/Cross-references: GDB:119802; OMIM:120650

A/Map position: 1q32-1q32

C/Superfamily: complement C3d/Epstein-Barr virus receptor; complement factor H repeat hom

C/Keywords: alternative splicing; duplication; glycoprotein; transmembrane protein

F/1-20/Domain: signal sequence #status predicted <SIG>

F/21-1091/Product: complement receptor 2 (16-repeat form) #status predicted <MAT>

F/121-658, 718-1091/Product: complement receptor 2 (15-repeat form) #status predicted <MAT>

F/23-82/Domain: complement factor H repeat homology <FHI>

F/91-146/Domain: complement factor H repeat homology <FHI>

Query Match 2.5%; Score 244; DB 1; Length 1091;  
Best Local Similarity 19.8%; Pred. No. 1.5e-07;  
Matches 138; Conservative 60; Mismatches 180; Indels 318; Gaps 37;

RESULT 8  
NBMSH

Query Match	2.4%	Score 238.5	DB 1	Length 1234
Best Local Similarity	17.8%	Pred. No. 4.2e-07		
Matches 144; Conservative	83;	Mismatches 239;	Indels	Gaps 35

QY 1181 IYTRKGYLDOWATRAYSSHEDKKCPVSLV--TGE-PSHSLICTSHRPDLNHRPLTGWFR 123

Db 453 IHDNGFLESSSIYALNRETSYRKQGYNTGTIGSGITCLQ-----NGWSP 501

**D**b****

```

QY 1238 ---CVASENETODDRSEOP--EGSLKKEDEVMLKVCENRGEARAF--IFLTDTGLVPG 1290
DB 502 QPSCIKS-----CDMPFENSITKNTRTWFLNKLDEYCELVGENEYKTKGSI-- 551
QY 1291 EHQQPVTLYLTVDRGNSHLGTYGLSCQ---HNPLINVTNHNQVLFHHTTSVLNF- 1345
DB 552 ----TCTYV-----GMSDTPSCERECBSVPLTRKLVSPRREK-----YRVGDLLEFS 596
QY 1346 ---SSPRGISAVAL-----RTSSRIGLSANSCISEDEQONHOGOSCIHRPCGQ 1393
DB 597 CHGSHVGPSPDSVOCYHNGMSPGPFTCKGQVASCAPPEILANGELNGAKKVEYSHGEVVKY 656
QY 1394 DSCPSLLDHDADVNTCT-----SIGGLMKCAI----- 1421
DB 657 DCKPRFLKGPKNKIQCVDMWMTLLPVCEEBERTCGDIPLEHSGAKSVPRYHNGDVEF 716
QY 1422 TCQRGRLAOLSSQOYIRPMQKEILLTSSGHDONVSCLYD---CGV-----PDP 1469
DB 717 ICEBFTMIGHGS-----VSCISGKWTQLPKCVATDLEKCRVLKSTGIEAIKP 765
QY 1470 SLVNYA-----NFSCEGTEFLKRCST-----SCVPAKLQGL----- 1502
DB 766 KLTFTFHTNSTMDYKCKRDKQYERSICINGKMDPEPNTCKTSCPPPOIENVOIETTVK 825
QY 1503 -----SPWLTCLDEGLM-SLPEVYCYLCECDAPRIILNANLLPHCL 1542
DB 826 YLDGEKLSVLCODNYLTVQSEBWC-KDGMQSLPRCTEKIPCSQPTIEHGSINLPRSS 884
QY 1543 QD-----NHVGTICKECKRPGYVVASAGKVRNKLKIQCLEGGIWEQSCIPV 1594
DB 885 EERRDSIESSSHGHTFTSYVCDGFRIPREN-----RLTCYMGKSTPBCVGLP 935
QY 1595 CERRPPY-----FEGMYECTNGFSL----- 1614
DB 936 CGEPPIPLCTVSLLESYQGEVYHGTGFGIDEPAPITCEGKMSPPKCIKTCD 995
QY 1615 ----- 1619
DB 996 VLPTVKNALIRGSKKSYRTGEOVTPRCQSPYQWNGSDYTVCVNSRMIGQPVCKDNCVD 1055
QY 1620 -----LNCQERE---KLPIICTKEGLMTQSPKICENIQGE 1652
DB 1056 PPHVPAIVTRTKNKKYHGDVRYECNKPELFGQVAVWC-ENGWTERPK-CRDSGK 1113
QY 1653 CRRPP-----SELNSVYKCEQGYGICA-----VCSPLCY 1682
DB 1114 CGEPPIPLDNGDITSLSPYVEPLUSSVBYQCKYLLKCKKTTCTNGKMSBPPTCLHACV 1173
QY 1683 IPSPDPVMLPENITADTLEHMEPEVKVQS 1711
DB 1174 IP--ENIMESHNT---ILKMRHTEKIYS 1196

RESULT 9
A35046
E-selectin precursor - human
N:Alternate names: endothelial leukocyte adhesion molecule 1 (ELAM-1); LECAM-2
C:Species: Homo sapiens (man)
C>Date: 27-Jul-1990 #sequence, revision 27-Jul-1990 #text, change 02-Aug-2002
C/Accession: A38615; A35046; A32606
R/Collins, T.; Williams, A.; Johnston, G.I.; Kilm, J.; Eddy, R.; Shows, T.; Gimbrone Jr., J. Biol. Chem. 266, 2466-2473, 1991
A>Title: Structure and chromosomal location of the gene for endothelial-leukocyte adhesion molecule 1
A/Reference number: A38615; MUID:9115870; PMID:11703529
A/Accession: A38615
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-610 <COL>
A:Cross-references: GB:M61893; GB:M58017; NID:g182043; PIDN:AAA52375.1; PID:g182046
R:Hesslein, C.; Osborn, U.; Golf, D.; Chi-Roseo, G.; Vassallo, C.; Pasek, M.; Piltack, C. Proc. Natl. Acad. Sci. U.S.A. 87, 1673-1677, 1990
A>Title: Endothelial leukocyte adhesion molecule 1: direct expression cloning and functi
A/Reference number: A35046; MUID:90175359; PMID:1689848

```

```

A/Accession: A35046
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-610 <HES>
A:Cross-references: GB:M30640; NID:g182047; PIDN:AAA52377.1; PID:g182048
R:Bevilacqua, M.P.; Stengelin, S.; Gimbrone Jr., M.A.; Seed, B. Science 243, 1160-1165, 1989
A>Title: Endothelial leukocyte adhesion molecule 1: an inducible receptor for neutrophils
A/Reference number: A32606; MUID:89162047; PMID:2466335
A/Accession: A32606
A/Molecule type: mRNA
A/Residues: 1-467, 469-610 <BEV>
A:Cross-references: GB:M24736; NID:g537523; PIDN:AAA52376.1; PID:g537524
C/Genetics:
A:Gene: GDB:SELB; ELAM; ESEL; ELAM1
A:Map position: 1q22-1q25
A:Superfamily: P-selectin, C-type lectin homology; complement factor H repeat homology; E
C/Keywords: duplication; glycoprotein; tandem repeat; transmembrane protein
F/1-21/Domain: signal sequence #status predicted <SIG>
F/12-13/Domain: C-type lectin homology <LCH>
F/22-610/Product: endothelial leukocyte adhesion molecule 1 #status predicted <MAT>
F/143-174/Domain: EGF homology <EGF>
F/180-237/Domain: complement factor H repeat homology <FH01>
F/242-299/Domain: complement factor H repeat homology <FH02>
F/304-362/Domain: complement factor H repeat homology <FH03>
F/367-425/Domain: complement factor H repeat homology <FH04>
F/430-488/Domain: complement factor H repeat homology <FH05>
F/493-547/Domain: complement factor H repeat homology <FH06>
F/557-578/Domain: transmembrane #status predicted <TM>
F/25,145,160,179,199,203,265,312,332,503,527/Binding site: carbohydrate (asn) (covalent)

```

```

Query Match 2.3%; Score 230.5; DB 2; Length 610;
Best Local Similarity 21.1%; Pred. No. 5.1e-07;
Matches 127; Conservative 80; Mismatches 215; Indels 181; Gaps 36;

QY 1230 RPLT-----GWPCVASENETODDRSEOPESGLKKEDEV--W-----LKYCFNRPGE 1274
DB 88 KPLTEBAKNMNP--GEPNNRQKD-EDCVELIYIRKDVGMNBERGSKKTLALCY----- 139
QY 1275 ARAIFLTDTGLVPEHQQPVTLYLTVDRGNSH-----SLGTY-----GLSCQH 1320
DB 140 -----TAACTNYSCHGHECVATINNYYTCKCPGFGSLKCEQ 176
QY 1321 NPILINVTNHNQVLFHHTTSVLNFPSPRGISAVALTSSRG-LSAPNRC-ISEDEGQ 1378
DB 177 --IVTCT-----ALSPERH-----SLVCSHPILGNFSYNSSCSISCRGY 214
QY 1379 NHOGSCIRHRCCKODSCPSLYLDHADVNTCTSI-----GPGIM-----KCAIT 1422
DB 215 LPSMETMQMMSGEMSAF--IPACNVVBCDAVTPNAPNPFVCEFGNPSFPNNTTCTFD 271
QY 1423 CQRGPAIQASSGOYIRPMQKEILLTSSGHW-ONVSCLEPVDCGV---PDPISLVNANFS 1478
DB 272 CEEGFEL-----MGAOSLQCTSSGMDNEKPKCAVTRAVRQPNQSVYRCSHP 321
QY 1479 CSEGTKEKRCSTISCVPRALQGLSPMLTCLDEGLMSLPEVYCK-LECDAPRIILNANL 1537
DB 322 AGEFT- FKSSCNFTCEGFMQLQPAQ-VBCTTGQWTOQIPVCEBAFCTA---LSNBERG 376
QY 1538 LPHCL---QDNHVGITCKECKRPGYVVASAGKVRNKLKIQCLEGGIW--EQQSCIP 1592
DB 377 YNNCLPSBAGSFRYSGSCERSCQGFYLSK-----RLQCGTGMENDEKPTCEA 427
QY 1593 VVCEP--PPVPEGMTVECTNG---PSLSQCVLNCQERE---KLPICTKEGLMTQEF 1643
DB 428 VRCDVAVHQP--KGLVRCAHSPIGEFYKSSCAFSCEBGEHLGSLQJLBCTSGQOWTEBV 485
QY 1644 KLCENIQGECPRRPPELSN-----VTKCEQGYGICA---VCS-----P 1679
DB 486 PSCQVYKCSLAVPGKINMSGCEPVGTVCKACBEGMTLNGSAARTCATGHWGSLP 545
QY 1680 LCVIPSPDPVMLPENITADTLE-HMMEPEVKVQIVCTGR-RQWHPDVLVNCIQSCERFQ 1737

```



Db 546 TCEAPTESNIFVLVAGLSAAGSLITLTPAFILMLRKCLRKAKKVPV-----ASSCSLE 598  
 QY 1738 ADG 1740  
 Db 599 SDG 601

## RESULT 10

146709  
 endothelial leukocyte adhesion molecule 1 - rabbit  
 C/Species: Oryctolagus cuniculus (domestic rabbit)  
 C/Date: 14-Feb-1997 #sequence\_revision 14-Feb-1997 #text\_change 16-Aug-2002  
 C/Accession: 146709; 146708  
 R/Larigan, J.D.; Tsang, T.C.; Rumberger, J.M.; Burns, D.K.  
 DNA Cell Biol. 11, 149-162, 1992  
 A/Title: Characterization of cDNA and genomic sequences encoding rabbit ELAM-1: Conserva  
 A/Reference number: 146708; MUID:92189729; PMID:1372169  
 A/Accession: 146709  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: mRNA  
 A/Residues: 1-551 <LARI>  
 A/Cross-references: GB:M91005; NID:g165006; PIDN:AAA31244.1; PID:g165007  
 A/Accession: 146708  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: mRNA  
 A/Residues: 1-307; 'T', 309-327; 'T', 329-490; 'A', 492-551 <LARI>  
 A/Cross-references: GB:M91004; NID:g165004; PIDN:AAA31243.1; PID:g165005  
 A/GeneticS:  
 A/Gene: ELAM1  
 A/Map position: 1q22-q25  
 A/Intons: 13/1; 143/1; 179/1; 241/1; 303/1; 366/1; 429/1; 488/1; 525/1; 533/2  
 C/Superfamily: P-selectin; C-type lectin homology; complement factor H repeat homology;  
 C/Keywords: duplication; glycoprotein; tandem repeat  
 F:14-140/Domain: C-type lectin homology <LCH>  
 F:162-239/Domain: complement factor H repeat homology <FH1>  
 F:244-301/Domain: complement factor H repeat homology <FH2>  
 F:306-364/Domain: complement factor H repeat homology <FH3>  
 F:369-427/Domain: complement factor H repeat homology <FH4>  
 F:432-486/Domain: complement factor H repeat homology <FH5>  
 F:32,45,201,314,321,466/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 2.3%; Score 228; DB 2; Length 551;  
 Best Local Similarity 20.0%; Pred. No. 6.4e-07;  
 Matches 105; Conservative 61; Mismatches 178; Indels 180; Gaps 26;  
 QY 1257 LKKEDEWLKCYKFRPGEARAIIFLTIDG---VPGF-----HQQPVITYL----- 1301  
 Db 76 IRKYNWVWVGTGHP-----LTEGAKNWAPEPNKKNNEDEVEIYIKRPKDTG 125  
 QY 1302 --TDVRGSNHSIG--TYGLSCQHNPLIIVTHNQVLFHHTTSVLLNFSSPRVGISAVALR 1358  
 Db 126 MNNDERCSKKLALCYTAC----- 145  
 QY 1359 TSSRIGLSAPNSCISEDEGQNHGOSCTHRP-----CGKODSCPSLLLDHADVNCST-S 1411  
 Db 146 --TASCSGHECI-----ETINNVSKCYGPFSGLKCEQVVTCEAQVQPHGSLNCHP 198  
 QY 1412 IG--PGLMKCAITCGRGFALDASSGQYIRPMQKEILITCSSGHND-QNVSLCPVDC--G 1465  
 Db 199 LGNFSYSSCSVSCERGL-----PSTETWTCTSSGSEAPRATCKVVECDTMG 248  
 QY 1466 VPDESLVYVNFSCSEGTGKFLKRCISICVPRAKLOGSLPW-----L 1506  
 Db 249 KP-----AN-----GDVKCSFS--QGSAPMNTTCTFDCEBGFLLGARSL 286  
 QY 1507 TCLEDGMSLEPVYCK--LECDAPRIILNANLLPHCLQDNHDVGTICKYCKPGYVAES 1565  
 Db 287 QCTSSGSDNEKRPCKAVSCDAIHNHPQSGSVSCNSSSGKFAFPASSCNFTCEENFLLRGP 346  
 QY 1566 AEGYVRKRLKIQCLBGGIWEQS--CIPIVCEPPRPVFEQMECTN--GSLDSQCV 1619  
 Db 347 AQ-----VECTAOGOMTOAPVCEAVKCDPVHTLEDGFVKTCTHPTGEFTYKSGCT 397

QY 1620 LNCNOERE---KLPICTKEGLWTOEFKLCENLQCECPPPELSNVEYKCEQYGIGAV 1676  
 Db 398 FNCRGEFELHGSQALECTSGQMAQELPSCQVVO--C-PSLAVYLGKTNVSCSGEPVGTV 454  
 QY 1677 CSPLCVIPSDPVMLPENITADTLEHMEFPVKQSVICTGRROW 1720  
 Db 455 CNFAC-----PEGWTLN-----GSAALMCGABGQW 479

## RESULT 11

153821  
 P-selectin - rat  
 C/Species: Rattus norvegicus (Norway rat)  
 C/Date: 29-May-1998 #sequence\_revision 29-May-1998 #text\_change 02-Aug-2002  
 C/Accession: 153821  
 R/Auchampach, J.A.; Oliver, M.G.; Anderson, D.C.; Manning, A.M.  
 Gene 145, 251-255, 1994  
 A/Title: Cloning, sequence comparison and in vivo expression of the gene encoding rat P-  
 A/Reference number: 153821; MUID:9433817; PMID:7520013  
 A/Accession: 153821  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: mRNA  
 A/Residues: 1-768 <RES>  
 A/Cross-references: GB:I23088; NID:g349552; PIDN:AAA60325.1; PID:g349553  
 C/Superfamily: P-selectin; C-type lectin homology; complement factor H repeat homology;  
 F:32-158/Domain: C-type lectin homology <LCH>  
 F:163-194/Domain: EGF homology <EGF>  
 F:262-319/Domain: complement factor H repeat homology <FHR>  
 F:510-567/Domain: complement factor H repeat homology <FHO6>  
 F:580-637/Domain: complement factor H repeat homology <FHO7>  
 F:642-699/Domain: complement factor H repeat homology <FHO8>

Query Match 2.3%; Score 227.5; DB 2; Length 768;  
 Best Local Similarity 20.4%; Pred. No. 1.1e-06;  
 Matches 105; Conservative 67; Mismatches 176; Indels 167; Gaps 28;  
 QY 1365 LSAPNSCISEDEGQNHGOSCIH-----RPGKQD-----SCPSLLLD 1402  
 Db 110 LTAENAEWADNEPNKKNNDCEVEIYIKNSAPQKNWDECFKRAKALCYTASQDM--- 166  
 QY 1403 HADVNCSTSIGP-----GLMKCAITCGRGFALDASSGQYIRPMQKEILITCSSGHNDQNV 1457  
 Db 167 -----SCNSQGERIETIGSYTC--SCYPGF--YGEPECEYV----- 198  
 QY 1458 SCLPVDG---VPRSLVYVNFSCSE---CTKFLKRCISICVPRALQGLSPWLTGLD 1511  
 Db 199 -----ECGKFDIPQHVLMN-----CSHPLADDFSSQCTFSCEPGYDLNGPSR-MQGLAS 247  
 QY 1512 GLMSLEPVYCK--LECDAPRIILNANLLPHCLQDNHDVGTICKYCKPGYVAESAGKV 1570  
 Db 248 GIWNNRPQCKAVQCSLEAPLHGTMDCTHPLA-AFAYDSCKFECPGY-----RM 298  
 QY 1571 RNKLKLIQCLBGGIWEQ--GSCIPVVCSEPPPVFEQMECTN--TNGSLDSQCVLNCNOE 1625  
 Db 299 RGSDI-LHCTIDSGQWBPPLTCEALACEPLESLPHGSMDCFPFGAGYSSCTFRCTEG 357  
 QY 1626 REKU---PIICTKRGTLTQEFKLCENLQCECPPRPSL-----NSVEKCE 1668  
 Db 358 FVLMGNDALHCADIGQWTAAPVCEALQCEFPVPSAQVSCSPDFPLKYQASCSFSCD 417  
 QY 1669 QG-----YGIGAVCSPLCVIPSDPVMLPENITADTLE----- 1701  
 Db 418 EGSLLVAGSVYRCLATGHSSEAPRECAVACSTPLSBNENGTMTCTQPLGHSNYKSTQGM 477  
 QY 1702 -----HMEFPVKQSVICTGRROWHPDVL--VHCIOSECFPGADGCDTINNR-----A 1749  
 Db 478 CDEGFYLSGERLD---CSPSGHWGTSRPMCEALICEPEIFAPBEGGSLDCSHVHGEFVGS 534  
 QY 1750 YCHVDGDCS---STSSKKVLPFADCDLDECT 1781  
 Db 535 TCHFS-----CNEBEFELGSRNV-----ECT 555



## RESULT 12

A35626

transforming growth factor beta-1-binding protein - human

C/Species: Homo sapiens (man)

C/Date: 21-Sep-1990 #sequence\_revision 21-Sep-1990 #text\_change 11-Jan-2000

C/Accession: A35626

R/Kanzaki, T.; Olofsson, A.; Moren, A.; Wernstedt, C.; Hellman, U.; Miyazono, K.; Claess

Cell 61, 1051-1061, 1990

A/Title: TGF-beta1 binding protein: a component of the large latent complex of TGF-beta1

A/Reference number: A35626; MUID:90275601; PMID:2350783

A/Accession: A35626

A/Status: preliminary; not compared with conceptual translation

A/Molecule type: mRNA

A/Residues: 1-1394 &lt;KAN&gt;

A/Cross-references: GB:M34057; NID:g339547; PIDN:AAA61160.1; PID:g339548

C/Superfamily: unassigned EGF-related proteins; EGF homology

C/Keywords: alternative splicing

F/750-791/Domain: EGF homology &lt;EGF&gt;

Query Match 2.3%; Score 225; DB 2; Length 1394;

Best Local Similarity 19.6%; Pred. No. 3.8e-06;

Matches 238; Conservative 121; Mismatches 379; Indels 476; Gaps 70;

QY 848 IPWVWIGQTKSLTITMLPISGVYDRAAGSLGACTEDGT---FRQYVH---TASS 899

DB 14 LPPLVSNHTRIRIVFTPSICKYTKK--GSCQNSCKKNTTLLISENGHADTLTATN 71

QY 900 RRV---CDSSGYWTPPEAVGPPVDVQCPSPSLQAWSPEVHLVHMMTVPCPTGCSL 953

DB 72 FRVIVCHLPCWNGQCSRDKC-----QCPNPTGKLCQI---FVHGASV 113

QY 954 ELLEQHHVQ-----ADTLTLMVTFPMESSQVLPDTILLNKESVHL-GLD 1000

DB 114 PKLYQHSQOPEKALGTHVHSTHTLPLVTS--QGVKVPKPPNIV---NIHVKHPPE 166

QY 1001 TFCDIPLTIKLVHDKVSGVAVYTFDERIEIDALITQSHPSLCSCGCPRYOVLDPP 1060

DB 167 A-----SVQIHQVSRIDGP-----TGQTKREKAPGOSQVSYO----- 198

QY 1061 FASGLPVVTVHS-HRKTFDVEVTRGOMYOYVLAEA-----GGELGEASPPIN-- 1107

DB 199 ---GLPVOKTQTIHSTYSHQVIR---HVPVAAKQUGRCFQETIGSQCKALPGLSKQ 252

QY 1108 -----HIKQ-----AAYCGDKVSE-RLGEEC 1128

DB 253 EDCCGVTGSMGFNKKCKCPKPSYHGNQMMECPLPGYKRVNNTFFCD--INECOLQGV 310

QY 1129 DDGDLVSGDGSKVCELEEGF-----NCVGPSPLCVMEGGDICEPFEKTSIVDCI 1181

DB 311 PNGECIAMTG-SYCTCKIGFGPPTPTSSCVDPDPV--ISEKGPC-----YRLVSSG 360

QY 1182 YTPKGYLDQWATRAYSSHEDKKCPVSLVTEBPSLIC-----TS 1221

DB 361 -----RQCMYPLSVHLTKQLCCSAYGKAGPHCEKRLPGTAFKEICGGMGYTVSG 412

QY 1222 YHPLDPMHRLPTGMFPCVASENETQD--DRSEQPSGLKKEDEWLKVCNPPGARAIFI 1280

DB 413 VHRRRPILHNHV-GGQPVFVKPKNTQPVAKSTHPPPLPKKEBPV-EALTFSEHGARSAP 470

QY 1281 FLFTDGLVPGEHQOQPTVLYLTQVSGNSHSLGTGLSCQHNPLILNTHQNVLFHTTS 1340

DB 471 EVAT---APPEKEIPLSDQETKLEPGQPLSP-GISAIH-----LHPQFP 512

QY 1341 VLNFPSSPPRGVIG-AVALRTSSRIGLAPNSCISEDE-----GQNHQOSCIHP-- 1389

DB 513 VVIEKTSPPVVEVAPEASTSSASQVLAFTQVTEINECTVAPDICGAGH---CINLPVR 568

QY 1390 -----CGNODSPSLLDHADVNTSTI-----GRMLKCATTCQGRF 1427

DB 569 YTCICVGYGRSEQORCK-----VIDICTOVQHLCSQRCENTGSSFLC--ICPAGF 619

QY 1428 ALQASSG-----QYIRPMQKEILLTCSGMDQNV----- 1457

## RESULT 13

T27283

hypothetical protein Y64G10A.f - Caenorhabditis elegans

C/Species: Caenorhabditis elegans

C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C/Accession: T27283

R/Ainscough, R.

submitted to the EMBL Data Library, September 1999

A/Reference number: Z20336

A/Accession: T27283

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Cross-references: EMBL:AL110498; NID:e1542303; PIDN:CAB54471.1; CESP:Y64G10A.f

A/Experimental source: clone Y64G10A

A/Genetic:

A/Introns: 77/1; 116/1; 198/1; 282/1; 365/1; 425/1; 466/1; 548/1; 559/1; 601/1; 625/1; 7

Query Match 2.3%; Score 224; DB 2; Length 1620;

Best Local Similarity 18.9%; Pred. No. 5.5e-06;

Matches 294; Conservative 125; Mismatches 457; Indels 682; Gaps 90;

QY 399 FMASCRSLILGDSBEDGH-----YFNGHGLTVLFW 429

DB 488 YMSCSEP---GFELSEGHSHCHDMNECLINNGCAOLCKNRKSGRRCCQCFAGYI----- 538

QY 430 STALPQSHFQHSQSHSGSEBEATDVLTLASFEPVNTMVPFRDEKYPRL---EVLOGFE 485

DB 539 -----LAHDEKSCVAASDAD-IFSNDIEDYS-----KVPGLDSIDEVISIE 580

QY 486 PEPILSLPRLPCGQVCNVELISQYNGWYPLRGKAVIRYQVNVICDBEGLNPIYSE 545

DB 581 STYADESP-RPLVFGRR--RHVRACVNFGTSLSE-----LFSSE 617

QY 546 QIRLOHEALNEAFSRVYISWOLSVHGVHNS---TLRRVVLVNGE-PSKIGNDHCECE 601

```

Db 618 VRTDSEKCPNGF--FGSTCOLSCSDCQNGKCSMRSGSLSKCDPCPSGYTEKCEQICR 675
Qy 602 HPLTGYDGGDRCLGRCTYSNRRROGLCHVECNMMLND--FDDG-----DC--- 644
Db 676 N---GYMGVDAHAKSCSLCDPSTGSCREDEKSDPCPDGFGYGCNLKCRWDPCNG 732
Qy 645 -CDPQVADVARTCPDPSPKRAYMSVKELKEALQJNSTHPL---NIYF-ASSVEDLAGA 699
Db 733 RCDP---VFQYCTCPDG-----LYGSCSEKPCPHFTFGKNCRCPCACARENSGC 779
Qy 700 ATWPMKDAVTHLGGIVLSPAYYMGPHGHTDMIEHGVHVLGLYHVFKEVSRSESCN--- 755
Db 780 -----DELTA--GKCRCKPGYTG--HHCKRMSP-----GLFGA--GCAMKSCSPAGIR 821
Qy 756 -DP---CKETVPSMENGDL-----CADTAPTPSELCREPEPISD 791
Db 822 CDPYTGDTCKKCPAGYQGNLCDQPCPAGYFGYDCEQKSCADVASPHKSKVCHH---VTG 878
Qy 792 TCGFTRPFG--APFTNMYSTDNCTDNFTNQVARMHCYLDLVYQWTESRKFTPIPI 849
Db 879 TC--TCLPGKTGPLIC-----DOC-----LIF----- 897
Qy 850 PMVIGQTNKSLTIHMLPISGVVYDRASGLCGACTEDGTRQYVHTAS--SRVYCDSSG 907
Db 898 -----VETIEFDIAFSINVIACAPTYGPNCAHTSCSYNGAKCDEBS- 938
Qy 908 YMTBEAVGPDPVQPCPSLQAMSPEVHALTHAMTVCPREGSLLELPHQVQADTLT 967
Db 939 -----DGSCHCTPG-----FGATCSEVCPT----- 959
Qy 968 LMTVTFMESQVLPDTEILLENKESVHLGPDFTCDIPLTKLHVDKSGVKYTFDE 1027
Db 960 ----- 960
Qy 1028 RIEIDALLTSQPHSPIC---SGCRPVRYOVLADPPASGLPVVYTHSRKFTDVEYTP 1083
Db 961 RFGIDCMQLCKCQNGAICDTNSGCECA-----PQMSG-----KKCDKACAP 1002
Qy 1084 GQMYOYOVLAAGSELGASPLPLNHIH-----GAPYCDGKVSERLGECDSDGLVSDG 1138
Db 1003 GTF-----GKDCSKCCDCAQGMHCDPSDDECTCPRGKHKHCDTCTCSG--LFGAG 1051
Qy 1139 CSKYCELEEGFNCGVEPSLCYMYEG-----DG-----ICEPERKTSIYDC 1179
Db 1052 CKGICSCQNGATCTSYNGSCRCRPGMRGKKCDRCPDRPREGCAIAD-----CTTNDT 1107
Qy 1180 GIYTPKGYLDQWATR-----AYSHEDKCKCPVSL-VTEPHSLICTSYHDL 1226
Db 1108 SMYNP-----FVARCDHVTEGRCRCPAGMTGPCQTSCLPLGRHSGGCRHSCQCSNG- 1157
Qy 1227 PNHRLPLGMFPCVASENTODRSEQRPBGLKKEDEVVLKYCFNRPGARAIPLFTLTDG 1286
Db 1158 ASCDRVTFGCDG--DSGFMGKNCSECPRG-----LM----- 1187
Qy 1287 LVPBEHQQPTVTLTLVDVRSN---HSLGTGYLSQONHPLIINYTHQNVLFHHTSVIL 1343
Db 1188 -----GSMCKHCKLCHMGGEENK----- 1206
Qy 1344 NFSSPRVGAISAVALTSSRIGLSAPNSCISEDQGN-----HOGSCIHRRPGCKODSC 1396
Db 1207 NGDECECID-----GWTGPSLCPFGQFGNCAQRCNCKKAGASC--DRKTGRCCCL 1253
Qy 1397 PSLLLDHADVNCISIGBGLMKCAITCQRFALQASSGOYIRPMQKELLTLCSSGHMDN 1456
Db 1354 PGMSGEHCE-KSCVS-GHYGAKCBETC-----ECENALCDPI-----SGH----- 1292
Qy 1457 VSCLP-----VDCGVDPSELVYVYANFSCSEGTFLKRCIS- 1503
Db 1293 CSCQPMRGKKCN--RPLCKIYFGRHCSQC-----RCANSGSCDHISRCQCPRGVAGHS 1346
Qy 1504 PMLTCLF---DGLWSLPEVYKLECDADPILIANLNLPLHCLQDNHVDVITCKYCKRGY 1560
Db 1347 -----CTELCPDGTG---ESCSQKCD-----CGENSMCDALISGKCFCKRGH 1385

```

```

Qy 1561 YVAESAEGYR-----NKLTKIQCLEGGIWDG--GSCIPYVCEPPPVPEGYECTNG 1611
Db 1386 SSSDKSGGVQVQGRFPPDNCQ---CSCENGVCDSSTSC--VC---PQYTG----- 1430
Qy 1612 FSLDQCVLNCOER-----EKPLICTKEGLWTOEFKLCENLOGECPPPELSNVEYK 1666
Db 1431 -----TKCEINACQSDRFPPTCEK---ICNCGNGT-----CDRLTQGC-----R 1466
Qy 1667 CEQGYIGAVCSPLCVIPPSDPVMLPENITADLTLEHMMEPYKVASIVCTGRQWHPDVL 1726
Db 1467 CLPGF-TGMYTCNOVC-----PE-----GR-----FG 1486
Qy 1727 VHCIOSECPQADQWCDTINNBAVCH--YDGDCCSSTLSKKYIPPAQDCDDECTC 1782
Db 1487 AGCKEKCR--CANHCNASSGCECKNIGFTGSPCEGSCPSGK-----YGLNCTLD-CEC 1537

```

## RESULT 14

```

A42755
P-selectin precursor - mouse
N/Alternate names: CD62; granule membrane protein 140; PADGEM
C/Species: Mus musculus (house mouse)
C/Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 16-Aug-2002
C/Accession: A42755; A44899
R/Weller, A.; Isenmann, S.; Vestweber, D.
J. Biol. Chem. 267, 15176-15183, 1992
A>Title: Cloning of the mouse endothelial selectin. Expression of both E- and P-selectin
A/Reference number: A42755; MUID:92340571; PMID:1378846
A/Accession: A42755
A>Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-768 <MW>
A/Cross-References: GB:M87861; NID:g200552; PIDN:AAA4008.1; PID:g200553
A/Experimental source: endothelial cells
A/Note: sequence extracted from NCBI backbone (NCBI:P109467)
R/Sanders, W.E.; Wilson, R.W.; Ballantyne, C.M.; Beaudet, A.L.
Blood 80, 795-800, 1992
A>Title: Molecular cloning and analysis of in vivo expression of murine P-selectin.
A/Reference number: A44899; MUID:92345617; PMID:1379089
A/Accession: A44899
A>Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-723, 'E', 725-768 <SN>
A/Cross-References: GB:M72332; NID:g193565; PIDN:AAA37712.1; PID:g193566
A/Note: sequence extracted from NCBI backbone (NCBI:P109900)
C/Superfamily: P-selectin; C-type lectin homology; complement factor H repeat homology;
C/Keywords: cell adhesion; glycoprotein; phospholipid; phosphoprotein; transmembrane
F/1-41/Domain: signal sequence #status predicted <SIG>
F/42-768/Product: P-selectin #status predicted <MAT>
F/161-194/Domain: EGF homology <EGF>
F/200-257/Domain: complement factor H repeat homology <FH01>
F/262-319/Domain: complement factor H repeat homology <FH02>
F/324-381/Domain: complement factor H repeat homology <FH03>
F/386-443/Domain: complement factor H repeat homology <FH04>
F/448-505/Domain: complement factor H repeat homology <FH05>
F/510-567/Domain: complement factor H repeat homology <FH06>
F/580-637/Domain: complement factor H repeat homology <FH07>
F/642-699/Domain: complement factor H repeat homology <FH08>
F/710-733/Domain: transmembrane #status predicted <TM>
F/734-768/Domain: intracellular #status predicted <INT>
F/45,54,107,212,347,398,456,467,603,654,661,679/Binding site: carbohydrate (asn) (covalent)

```

Query Match 2.28; Score 219; DB 2; Length 768;

Best Local Similarity 23.11; Pred. No. 4e-06;

Matches 99; Conservative 38; Mismatches 146; Indels 146; Gaps 29;

```

Qy 1384 SCIRPCKGKDCPSLLLDHADVNCISIGBGLMKCAITCQRFALQASSGOYIRPMQK- 1442
Db 162 SCQDMSCSNQCELEETI-----GSTTC--SCYPGF--YGEPEYVKEGKV 203
Qy 1443 -----EILLTCS--GHMDQVNSCLPVDGVPDPSLVYVYANFSCSEGTFLKRCISCVPP 1496

```



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2004, 16:02:37 ; Search time 59 Seconds  
(without alignments)  
7833.427 Million cell updates/sec

Title: US-09-983-025a-2  
Perfect score: 9856  
Sequence: 1 MMCKLRISLIALAGWALC.....AADCDLDECTCRDPKAEENQ 1791

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues  
Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_23:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriaph:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9856	100.0	1791	4 Q9BXP8	Q9BXP8 homo sapien
2	9831	99.7	1790	4 Q96P88	Q96P88 homo sapien
3	8945	90.8	1624	4 Q9H4C9	Q9H4C9 homo sapien
4	5281.5	53.6	1214	11 Q8BAG6	Q8BAG6 mus musculu
5	4394	44.6	826	4 Q96PH7	Q96PH7 homo sapien
6	3974	40.3	1545	11 Q8RAK8	Q8RAK8 mus musculu
7	3949.5	40.1	1574	11 Q8RAK7	Q8RAK7 mus musculu
8	3517	35.7	1367	11 Q9ES06	Q9ES06 mus musculu
9	2173	22.0	219	4 Q9NUP4	Q9NUP4 homo sapien
10	1255	12.7	468	11 Q9UK57	Q9UK57 homo sapien
11	1102.5	11.2	468	11 Q9UK57	Q9UK57 mus musculu
12	896.5	9.1	354	11 Q8K423	Q8K423 mus musculu
13	813.5	8.3	283	6 Q9SL43	Q9SL43 bos taurus
14	752.5	7.6	246	6 Q9SL44	Q9SL44 ovis aries
15	635.5	6.4	213	6 Q9SL42	Q9SL42 sus scrofa
16	625.5	6.3	212	6 Q9SL41	Q9SL41 equus caball

17	346.5	3.5	3567	11 Q9ES77	Q9ES77 mus musculu
18	292.5	3.0	2489	4 Q16744	Q16744 homo sapien
19	290	2.9	3564	11 Q923L3	Q923L3 mus musculu
20	289	2.9	1014	6 Q29530	Q29530 pan troglod
21	282.5	2.9	1574	11 Q88281	Q88281 ratu
22	281	2.9	1497	4 Q8NBT9	Q8NBT9 homo sapien
23	276	2.8	1911	6 Q29528	Q29528 papio hamad
24	271	2.7	669	11 Q922H0	Q922H0 mus musculu
25	270	2.7	601	11 Q9CUT3	Q9CUT3 mus musculu
26	259	2.6	1236	11 Q91YB6	Q91YB6 ratu
27	252	2.6	790	4 Q8NIE9	Q8NIE9 homo sapien
28	252	2.6	830	4 Q81VD1	Q81VD1 homo sapien
29	249.5	2.5	3508	4 Q96RM4	Q96RM4 homo sapien
30	247	2.5	610	6 Q95LG1	Q95LG1 equus caball
31	247	2.5	3389	4 Q96C09	Q96C09 homo sapien
32	236	2.4	1045	6 Q45545	Q45545 ovis aries
33	236	2.4	1637	6 Q9XSV8	Q9XSV8 bos taurus
34	236	2.4	5146	6 Q8SPM4	Q8SPM4 bos taurus
35	229.5	2.3	740	4 Q95508	Q95508 homo sapien
36	229.5	2.3	609	6 Q9GLF0	Q9GLF0 canis fami
37	229.5	2.3	1664	5 Q9TV02	Q9TV02 caenorhabdi
38	223	2.3	740	4 Q95507	Q95507 homo sapien
39	222	2.3	1234	6 Q8M172	Q8M172 sus scrofa
40	221.5	2.2	3857	11 Q88840	Q88840 mus musculu
41	219.5	2.2	1713	11 Q88349	Q88349 mus musculu
42	215.5	2.2	1713	11 Q8CG19	Q8CG19 mus musculu
43	215	2.2	2872	11 Q9WTH8	Q9WTH8 ratu
44	214.5	2.2	1032	11 Q9DC83	Q9DC83 mus musculu
45	213.5	2.2	343	6 Q9MY19	Q9MY19 erythrocebu

## ALIGNMENTS

RESULT 1	
Q9BXP8	PRELIMINARY; PRT; 1791 AA.
ID Q9BXP8	
AC Q9EXPE8	
DT 01-JUN-2001 (TREMUR1.17, Created)	
DT 01-JUN-2001 (TREMUR1.17, Last sequence update)	
DT 01-MAR-2003 (TREMUR1.23, Last annotation update)	
DE Pregnancy-associated Plasma preproprotein-A2.	
OS Homo sapiens (Human)	
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.	
OC NCBI_TaxID=9606;	
RN [1]	
RP SEQUENCE FROM N.A.	
RC TISSUE=Term placenta;	
RX MEDLINE=21293000; Pubmed=11264294;	
RA Overgaard M.T., Boldt H.B., Laurson L.S., Sottrup-Jensen L.,	
RA Conover C.A., Oxvig C.,	
RT "Pregnancy-associated Plasma Protein-A2 (PAPP-A2), a Novel Insulin-	
RT like Growth Factor-binding Protein-5 Proteinase.";	
RL J. Biol. Chem. 276:21849-21853(2001).	
DR EMBL: AF311940; AK31073.1; --	
DR MEROPS; M46.002; --	
DR Genew; HGNC:14615; PLAC3.	
DR InterPro; IPR003961; FN III.	
DR InterPro; IPR002052; N6 Itae.	
DR InterPro; IPR000800; Notch.	
DR InterPro; IPR000435; Sushi_SCR_CCP.	
DR InterPro; IPR006025; Zn_MTpeptidase.	
DR Pfam; PF00084; sushi; 4.	
DR SMART; SM00032; CCP; 4.	
DR SMART; SM00060; FN3; 1.	
DR SMART; SM00004; NL; 2.	
DR PROSITE; PS00092; N6_MTASE; 1.	
DR PROSITE; PS00142; ZINC_PROTEASE; 1.	
SQ SEQUENCE 1791 AA; 198537 MW; F436030821ECEDD CRC64;	
Query Match	100.0%; Score 9856; DB 4; Length 1791;
Best local Similarity	100.0%; Pred. No. 0;

Matches 1791; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Qy	1	MMCKLIRISLAIAGMALGSANSELGWTREKSLVEREHLNOVLEGERCMLGAKVRPR	60
Db	1	MMCKLIRISLAIAGMALGSANSELGWTREKSLVEREHLNOVLEGERCMLGAKVRPR	60
Qy	61	ASPOHLLFGVYPSRAGNYLREYPVGEDEIHHTGSRKPDTEGNANVSLVPPDLTENPAGLRG	120
Db	61	ASPOHLLFGVYPSRAGNYLREYPVGEDEIHHTGSRKPDTEGNANVSLVPPDLTENPAGLRG	120
Qy	121	AVEBPAPWVGDSPRIGSELLGDDDAYIGNORSKESLCEBAGIQKGSAMAATTTAIFPTL	180
Db	121	AVEBPAPWVGDSPRIGSELLGDDDAYIGNORSKESLCEBAGIQKGSAMAATTTAIFPTL	180
Qy	181	NEPPEIORGRGMASSROKROYWKRAEDQGDGSGISHFOPMKSLGHRVKSPPESN	240
Db	181	NEPPEIORGRGMASSROKROYWKRAEDQGDGSGISHFOPMKSLGHRVKSPPESN	240
Qy	241	QNGEGSYREAEFTNSQVGLPILYFSGRRELLRPEVLAIEIPREAFTEAMVXPEGON	300
Db	241	QNGEGSYREAEFTNSQVGLPILYFSGRRELLRPEVLAIEIPREAFTEAMVXPEGON	300
Qy	301	NPALIAAGVFDNCSTVSKGVALGIRSGDKGKRDARFFSLCTDRVKKATILISHSRVQ	360
Db	301	NPALIAAGVFDNCSTVSKGVALGIRSGDKGKRDARFFSLCTDRVKKATILISHSRVQ	360
Qy	361	PGTTHVATYDGRHMALYVDGTOVASLDSQSPNSPFMASCSLLIGDSEDEGHR	420
Db	361	PGTTHVATYDGRHMALYVDGTOVASLDSQSPNSPFMASCSLLIGDSEDEGHR	420
Qy	421	GHGLTVWSTALPOSHFOHSSOHSSGEEBATDLVLTASFEPTVTEWVPFDEKXPRLV	480
Db	421	GHGLTVWSTALPOSHFOHSSOHSSGEEBATDLVLTASFEPTVTEWVPFDEKXPRLV	480
Qy	481	LOGEPEPEELISPLQPLCGQTVCDNVELISQYNGWPLRSGKATIRYVANI	540
Db	481	LOGEPEPEELISPLQPLCGQTVCDNVELISQYNGWPLRSGKATIRYVANI	540
Qy	541	IVSEQIQLQHEBALNEAFSRYNISQWLSHOVHNSLHRYVLVNCEPSKIGNDHCEDEC	600
Db	541	IVSEQIQLQHEBALNEAFSRYNISQWLSHOVHNSLHRYVLVNCEPSKIGNDHCEDEC	600
Qy	601	EHPITGVYDGDGCRLOGRCYSMNRDGLCHEVCNNMLNFDGDCDDPOVADVRKTCFPPD	660
Db	601	EHPITGVYDGDGCRLOGRCYSMNRDGLCHEVCNNMLNFDGDCDDPOVADVRKTCFPPD	660
Qy	661	SPKRAYMSVKEALQJNSTHFLNIYFASVREDLAGAATPMDKDAVTHLGGIVLSPA	720
Db	661	SPKRAYMSVKEALQJNSTHFLNIYFASVREDLAGAATPMDKDAVTHLGGIVLSPA	720
Qy	721	YXGMPGHDTMIHEGVHGLYHVFKGVSERESCNDDPKETVPSMETGDLCAADTAAPPKS	780
Db	721	YXGMPGHDTMIHEGVHGLYHVFKGVSERESCNDDPKETVPSMETGDLCAADTAAPPKS	780
Qy	781	ELCBEPETSDTCGTFRPGAPFTNYMSYTDNCTDNTPNOVAMHCYLDLVYQOMTES	840
Db	781	ELCBEPETSDTCGTFRPGAPFTNYMSYTDNCTDNTPNOVAMHCYLDLVYQOMTES	840
Qy	841	RKPTPIPIPPWVIGQTNKSLTIHMLPRLISGVYVYDASGLCGACTEDGTFFQYVHTASSR	900
Db	841	RKPTPIPIPPWVIGQTNKSLTIHMLPRLISGVYVYDASGLCGACTEDGTFFQYVHTASSR	900
Qy	901	RVCSSGSGWPEBEAAGPPDVOPCPSLQAMSPEVHLVHMNTVPCPEGSGLELFPHP	960
Db	901	RVCSSGSGWPEBEAAGPPDVOPCPSLQAMSPEVHLVHMNTVPCPEGSGLELFPHP	960
Qy	961	VOADTLTLMTVSFFMESSQVLFTEILLENKESVHLAGLDTFCODIPLTIKHLVNDKVS	1020
Db	961	VOADTLTLMTVSFFMESSQVLFTEILLENKESVHLAGLDTFCODIPLTIKHLVNDKVS	1020
Qy	1021	KVYTFDERIEIDALLISQHPSLCSGCRPVRYQVLDPPASGLPVVVTSHRKTVE	1080
Db	1021	KVYTFDERIEIDALLISQHPSLCSGCRPVRYQVLDPPASGLPVVVTSHRKTVE	1080

Qy	1081	VTPCOMTQOVLAAGELGASAPPLNHIHGAAPYCGDKYSERLGEBCDDGDLVSGDGS	1140
Db	1081	VTPCOMTQOVLAAGELGASAPPLNHIHGAAPYCGDKYSERLGEBCDDGDLVSGDGS	1140
Qy	1141	KVCELEBGNVCVGPSPSLCYMYEGDICEPPEKRTSIVDCGIYTPKGYLDQMATRAYSSHE	1200
Db	1141	KVCELEBGNVCVGPSPSLCYMYEGDICEPPEKRTSIVDCGIYTPKGYLDQMATRAYSSHE	1200
Qy	1201	DKKCCPVSLVTGEPHSLICTSYNHPDLPHNRPLTGMPCVASENETODDRSEQPEGSLKE	1260
Db	1201	DKKCCPVSLVTGEPHSLICTSYNHPDLPHNRPLTGMPCVASENETODDRSEQPEGSLKE	1260
Qy	1261	DEVWLKCFNRPGEARAFITFLTTDGLVPEHQOPTYTLVTFVRGSNHSIGTYGLSCQH	1320
Db	1261	DEVWLKCFNRPGEARAFITFLTTDGLVPEHQOPTYTLVTFVRGSNHSIGTYGLSCQH	1320
Qy	1321	NPLIINTVTHQNVLFHHTTSVLLNFSSPRVGSISAVALTSSRIQLSAPSNCSIEDEGQNH	1380
Db	1321	NPLIINTVTHQNVLFHHTTSVLLNFSSPRVGSISAVALTSSRIQLSAPSNCSIEDEGQNH	1380
Qy	1381	QGOSCHRPCKGKDCSLLDHDADVNCISIGRLMKCAITQORGFALQASSGOYIRPM	1440
Db	1381	QGOSCHRPCKGKDCSLLDHDADVNCISIGRLMKCAITQORGFALQASSGOYIRPM	1440
Qy	1441	QKEILLTCSGSHMDQVSCLPVDCGVDPRLVNVANFSCSEGTFTLKRCSISCVPAKLQ	1500
Db	1441	QKEILLTCSGSHMDQVSCLPVDCGVDPRLVNVANFSCSEGTFTLKRCSISCVPAKLQ	1500
Qy	1501	GLSPWLTCLEBDGLMSLPEVYCKLECDAPITILNANILLPHCLDNDHVGITCKYECKPY	1560
Db	1501	GLSPWLTCLEBDGLMSLPEVYCKLECDAPITILNANILLPHCLDNDHVGITCKYECKPY	1560
Qy	1561	YVASSAGKRNKLLKIQCLEGGIWEQSGCIPVACEPPVFGMECTNGFSLDSQCVL	1620
Db	1561	YVASSAGKRNKLLKIQCLEGGIWEQSGCIPVACEPPVFGMECTNGFSLDSQCVL	1620
Qy	1621	NCNORERKLPILCTKEGLMTQEFKLCENLQGECPPEPSELSNVEYKCEQGYIGAVCSPL	1680
Db	1621	NCNORERKLPILCTKEGLMTQEFKLCENLQGECPPEPSELSNVEYKCEQGYIGAVCSPL	1680
Qy	1681	CVIPSPDPMVLPENITADTLEHMMEPVKQSIYCTGRQWHPVLAHICQSECFQADG	1740
Db	1681	CVIPSPDPMVLPENITADTLEHMMEPVKQSIYCTGRQWHPVLAHICQSECFQADG	1740
Qy	1741	WCDTINNRAYCHVDGCGCCSTLSSKVIIPPAADCDLDECTCRDPKAEHQ	1791
Db	1741	WCDTINNRAYCHVDGCGCCSTLSSKVIIPPAADCDLDECTCRDPKAEHQ	1791

## RESULT 2

Q96PH8 PRELIMINARY; PRT: 1790 AA.

AC Q96PH8; PT 01-DEC-2001 (Tremblrel. 19, Created)

DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)

DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)

DE Pregnancy-associated plasma protein E1.

GN PAPPE.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

NCBI\_TaxId=9606;

EN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=2148266; PubMed=11597188;

RA Page N.M., Butlin D.J., Lomthaisong K., Lowry P.J.;

RT "The Characterization of Pregnancy Associated Plasma Protein-E and the

RT Identification of an alternative Splice Variant.",

KL Placenta 22:681-687(2001).

DR EMBL; AF342989; A011779.1; -

DR InterPro; IPR002052; N6\_Mcase.

DR InterPro; IPR000800; Notch.

DR InterPro: IPR000436; Sush1 SCR CCP.  
DR InterPro: IPR006025; Zn\_MTPpeptidase.  
DR Pfam: PF000084; sush1\_4.  
DR SMART: SM00032; CCP; 4.  
DR SMART: SM00004; NL; 2.  
DR PROSITE: PS00092; N6\_MTASE; 1.  
DR PROSITE: PS00142; ZINC\_PROTASE; 1.  
SQ SEQUENCE 1790 AA; 198450 MW; E80717B6E623A0E3 CACC64;

Query Match	99.7%	Score 9931	DB 4	Length 1790
Best Local Similarity	99.8%	Pred. No. 0		
Matches 1787	Conservative	0	Mismatches 3	Indels 0
Gap				
0				
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				
21				
22				
23				
24				
25				
26				
27				
28				
29				
30				
31				
32				
33				
34				
35				
36				
37				
38				
39				
40				
41				
42				
43				
44				
45				
46				
47				
48				
49				
50				
51				
52				
53				
54				
55				
56				
57				
58				
59				
60				
61				
62				
63				
64				
65				
66				
67				
68				
69				
70				
71				
72				
73				
74				
75				
76				
77				
78				
79				
80				
81				
82				
83				
84				
85				
86				
87				
88				
89				
90				
91				
92				
93				
94				
95				
96				
97				
98				
99				
100				

QY	902	VCDSGGWTEBEAAGPDDVOPCPSPLOAASPEVHLVHNMMVYCPCEGSGLELLFQHPV	964
Db	901	VCDSGGWTEBEAAGPDDVOPCPSPLOAASPEVHLVHNMMVYCPCEGSGLELLFQHPV	960
QY	962	QADTLTLMWVSFFMESSQVLPDETEILLENNESVHLGPDTPCDIPLTIKLVHGKVSVK	1021
Db	961	QADTLTLMWVSFFMESSQVLPDETEILLENNESVHLGPDTPCDIPLTIKLVHGKVSVK	1020
QY	1022	VYTTDESEIEIDALLTSQPHSPILCSGGRPVRYOVLARDPPASGLPVVYTHSHKFTDVEV	1081
Db	1021	VYTTDESEIEIDALLTSQPHSPILCSGGRPVRYOVLARDPPASGLPVVYTHSHKFTDVEV	1080
QY	1082	TPGGMVYOVLAAAGSLGEASPLVNIHICAPYCGDGKVSBERLGEEDDDIVSGDCSK	1141
Db	1081	TPGGMVYOVLAAAGSLGEASPLVNIHICAPYCGDGKVSBERLGEEDDDIVSGDCSK	1140
QY	1142	VCELESEFNCVGEBSLTCYMEGDGICERPERKTSIVDCGITYTPKGYLDQWATRAYSSHD	1201
Db	1141	VCELESEFNCVGEBSLTCYMEGDGICERPERKTSIVDCGITYTPKGYLDQWATRAYSSHD	1200
QY	1202	KKCPVSLVTGEPHSLICTSYHPDLPHMRPLTGMFCVASENETODDRSQEGSLKED	1261
Db	1201	KKCPVSLVTGEPHSLICTSYHPDLPHMRPLTGMFCVASENETODDRSQEGSLKED	1260
QY	1262	EWMLKVCFNRGEARAFIFLITDGLVPGHQPTTLYLTVDRGSHSLGTYSLSQHN	1321
Db	1261	EWMLKVCFNRGEARAFIFLITDGLVPGHQPTTLYLTVDRGSHSLGTYSLSQHN	1320
QY	1322	PLIINTVHHQNVLPHTTSLVLPNPSRPVIGSAVALTSRIGASAPNSCISEDEGQNH	1381
Db	1321	PLIINTVHHQNVLPHTTSLVLPNPSRPVIGSAVALTSRIGASAPNSCISEDEGQNH	1380
QY	1382	GQSCIHRRPCGKODSCPSLLDHDADVNCISIGPLMKCAITCQGFALQASSGQYIRPMQ	1441
Db	1381	GQSCIHRRPCGKODSCPSLLDHDADVNCISIGPLMKCAITCQGFALQASSGQYIRPMQ	1440
QY	1442	KEILLTSSSGHMDQNVCLPVDGVPBPISLVNANBSCSEGTAKLKXCSISCVAPALQOG	1501
Db	1441	KEILLTSSSGHMDQNVCLPVDGVPBPISLVNANBSCSEGTAKLKXCSISCVAPALQOG	1500
QY	1502	LSPLWLTCELDGMSLPVYCKLECDAPITILNANLLPHCLQDNHVDGTCIKYCKRGYY	1561
Db	1501	LSPLWLTCELDGMSLPVYCKLECDAPITILNANLLPHCLQDNHVDGTCIKYCKRGYY	1560
QY	1562	VASBAEKGVNKLKICQLEBGIMWQSCIPVVCBEPPEPVEGMYECTNGFSLDSQCVLN	1621
Db	1561	VASBAEKGVNKLKICQLEBGIMWQSCIPVVCBEPPEPVEGMYECTNGFSLDSQCVLN	1620
QY	1622	CNOEREXPLITCTYEGIMTQEPKICENLQCBECPPPELSNVSEYKCEQYIGAVCSPLC	1681
Db	1621	CNOEREXPLITCTYEGIMTQEPKICENLQCBECPPPELSNVSEYKCEQYIGAVCSPLC	1680
QY	1682	VIPSPSDVMLPENTADTLTLEHMEPVVQSVICVGRQWMPDPLVNICIOSCEPQADGW	1741
Db	1681	VIPSPSDVMLPENTADTLTLEHMEPVVQSVICVGRQWMPDPLVNICIOSCEPQADGW	1740
QY	1742	CDTINNRAAYCHYDGDCCSSTLASKEYIPAAADCDIDECTCRDPKAEENQ	1791
Db	1741	CDTINNRAAYCHYDGDCCSSTLASKEYIPAAADCDIDECTCRDPKAEENQ	1790

RESULT 3

Q9H4C9 PRELIMINARY; PRT; 1624 AA. 1022

Q9H4C9

01-MAR-2001 (TIREMBLrel. 16, Created)

01-MAR-2001 (TIREMBLrel. 16, Last sequence update)

01-MAR-2003 (TIREMBLrel. 23, Last annotation update)

Pregnancy-associated plasma protein-E.

GN PAPB.

OS Homo sapiens (Human).

OC Buxatryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI\_TaxID=9606;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RX TISSUE=Placenta;  
 RA MEDLINE=20472054; PubMed=11018262;  
 RA Fair M., Stride J., Geppert H.G., Kocorek A., Mahne M., Tesche H.;  
 RL "Pregnancy-associated plasma protein-E (PAPP-E).";  
 RL Biochim. Biophys. Acta 1493:356-362(2000).  
 DR EMBL: AJ278348; CAC1134.1; -  
 DR MEROPS: M46.002; -  
 DR InterPro: IPR003961; FN III.  
 DR InterPro: IPR02052; N6\_Mcase.  
 DR InterPro: IPR00800; Notch.  
 DR InterPro: IPR00436; Sushi\_SCR CCP.  
 DR InterPro: IPR006025; Zn\_Mtpeptidase.  
 DR Pfam: PF00084; sushi; 4.  
 DR SMART: SM00032; CCP; 4.  
 DR SMART: SM00060; FN3; 1.  
 DR SMART: SM00004; NL; 2.  
 DR PROSITE: PS00092; N6\_MTASE; 1.  
 DR PROSITE: PS00142; ZINC\_PROTEASE; 1.  
 SQ SEQUENCE 1624 AA; 180426 MW; 106E2F1FC9CB2CB5 CRC64;

Query Match 90.8%; Score 8945; DB 4; Length 1624;  
 Best Local Similarity 99.8%; Pred. No. 0;  
 Matches 1620; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 168 MAATTTTAAFTTLNPKETQRGAKSRORROWKRAEDGGDSGSHHPMPKXSL 227  
 DB 1 MAATTTTAAFTTLNPKETQRGAKSRORROWKRAEDGGDSGSHHPMPKXSL 60  
 QY 228 KHRVKSPPESNONGEGSYREAEFTFNSQVGLPIYFSGRRRLRLPEVLAIPREAF 287  
 DB 61 KHRVKSPPESNONGEGSYREAEFTFNSQVGLPIYFSGRRRLRLPEVLAIPREAF 120  
 QY 288 TVEAWVKEGGQNNPAIIAGVFDNCSHTVSDKGMALGIRSGDKGRDARFPFSLCTDRV 347  
 DB 121 TVEAWVKEGGQNNPAIIAGVFDNCSHTVSDKGMALGIRSGDKGRDARFPFSLCTDRV 180  
 QY 348 KKATLLLSHSRYOQPTWTHVAATYDGRHMAIYVQCTVASSLDSGRLNSPFMASCRSL 407  
 DB 181 KKATLLLSHSRYOQPTWTHVAATYDGRHMAIYVQCTVASSLDSGRLNSPFMASCRSL 240  
 QY 408 LGGDSSEDEGHFRGLGLTVFWSTALPQSHFQHSQSSGGERATDVLTAPEPVNTW 467  
 DB 241 LGGDSSEDEGHFRGLGLTVFWSTALPQSHFQHSQSSGGERATDVLTAPEPVNTW 300  
 QY 468 VPFDEKYPRLVLTQGFPEPEIISPLQPLCGQTVCDNVELISQYNGWPLRGEKVIKY 527  
 DB 301 VPFDEKYPRLVLTQGFPEPEIISPLQPLCGQTVCDNVELISQYNGWPLRGEKVIKY 360  
 QY 528 QVANNICDEGNLPYVSEQIRLQHEALNEAFSRNYISQLSVHQVHNSTLHRRVYLVNCE 587  
 DB 361 QVANNICDEGNLPYVSEQIRLQHEALNEAFSRNYISQLSVHQVHNSTLHRRVYLVNCE 420  
 QY 588 PSKIGNDHCEDECEHPTLGYDGGDCRLQGRCSWNRDGLCHECNMNLNFDGDDCCDP 647  
 DB 421 PSKIGNDHCEDECEHPTLGYDGGDCRLQGRCSWNRDGLCHECNMNLNFDGDDCCDP 480  
 QY 648 QVADVKTCTFDPDSFKRAYMSVKELKEALQINSTHFLNIYFASVREDLAGAATWPMKD 707  
 DB 481 QVADVKTCTFDPDSFKRAYMSVKELKEALQINSTHFLNIYFASVREDLAGAATWPMKD 540  
 QY 708 AVTHLGGIVLSPAYYGMGHTDTMHEVGLGLYHVKVGSERSSCNDPCKEYTPSMET 767  
 DB 541 AVTHLGGIVLSPAYYGMGHTDTMHEVGLGLYHVKVGSERSSCNDPCKEYTPSMET 600  
 QY 768 GDLCAADTAPTPKSELCREPEPTSDTCGTRFPAGAFITYMSTYTDNCTDNFTPNOVAMH 827  
 DB 601 GDLCAADTAPTPKSELCREPEPTSDTCGTRFPAGAFITYMSTYTDNCTDNFTPNOVAMH 660  
 QY 828 CYLDLVYQWMTESRRKPTPIPPVNIQGTNKSJTIHMLPPISGVVDASGSLCGACTED 887

DB 661 CYLDLVYQWMTESRRKPTPIPPVNIQGTNKSJTIHMLPPISGVVDASGSLCGACTED 720  
 QY 888 GTFROYVHTASSRRVCSGWTPEEAVGPPVDQCEPSLQAMSEVHLXHMNTVPCP 947  
 DB 721 GTFROYVHTASSRRVCSGWTPEEAVGPPVDQCEPSLQAMSEVHLXHMNTVPCP 780  
 QY 948 TEGCSLELRFQHPVQADTLTLWTSFPMSSQVLPTEILLENKESVHLGLPDTFCDIPL 1007  
 DB 781 TEGCSLELRFQHPVQADTLTLWTSFPMSSQVLPTEILLENKESVHLGLPDTFCDIPL 840  
 QY 1008 TIKLHNDGKXSVKYVYTFEDRIRIIDAALLTSQHSPLCSGCCRVRVQVLRDPPASGLPV 1067  
 DB 841 TIKLHNDGKXSVKYVYTFEDRIRIIDAALLTSQHSPLCSGCCRVRVQVLRDPPASGLPV 900  
 QY 1068 VVTHSHRKFTDVEVTPQOMTQVYVLAAGELAEASPLNHIIGAPICGDKGKSERLGE 1127  
 DB 901 VVTHSHRKFTDVEVTPQOMTQVYVLAAGELAEASPLNHIIGAPICGDKGKSERLGE 960  
 QY 1128 CDDGDLVSGDCSKVCELEBGFNCVGEPSLCYMYEGDICEPERKTSIVDCGIYTPKGY 1187  
 DB 961 CDDGDLVSGDCSKVCELEBGFNCVGEPSLCYMYEGDICEPERKTSIVDCGIYTPKGY 1020  
 QY 1188 LDQWATRAYSHEDKKKCPVSLVTGEPHSLICTSYHPDLFNHRPLTGMPFCVASENETOD 1247  
 DB 1021 LDQWATRAYSHEDKKKCPVSLVTGEPHSLICTSYHPDLFNHRPLTGMPFCVASENETOD 1080  
 QY 1248 DRSEQPEGLKKEDEVWLKVCFNRPGEARAIFFLTDDGLVPGEHOOPTVLYLTVRGS 1307  
 DB 1081 DRSEQPEGLKKEDEVWLKVCFNRPGEARAIFFLTDDGLVPGEHOOPTVLYLTVRGS 1140  
 QY 1308 NMSLGTGYGLSCQNNPLIINVTTHQNVLFHHTSVLNFSSPRVIGISAVALRTSSRIGLSA 1367  
 DB 1141 NMSLGTGYGLSCQNNPLIINVTTHQNVLFHHTSVLNFSSPRVIGISAVALRTSSRIGLSA 1200  
 QY 1368 PNCISIDEQONHQGSCCHRPCKGKODSCPSLLDHADVNCISIGGLMKCAITQCRGF 1427  
 DB 1201 PNCISIDEQONHQGSCCHRPCKGKODSCPSLLDHADVNCISIGGLMKCAITQCRGF 1260  
 QY 1428 ALQASSGOYIRPMOKELILTCSSGHMDQVNSCLPVDGVPDPILVNYANFSCSEGTFLK 1487  
 DB 1261 ALQASSGOYIRPMOKELILTCSSGHMDQVNSCLPVDGVPDPILVNYANFSCSEGTFLK 1320  
 QY 1488 RCSIISCVPAKLOGLSPMLTCLBDGLMSLREYVCKLECDAPPIILNANLLPRLCLQDNHD 1547  
 DB 1321 RCSIISCVPAKLOGLSPMLTCLBDGLMSLREYVCKLECDAPPIILNANLLPRLCLQDNHD 1380  
 QY 1548 VGTICKECKRPGYVVAESAGKVRNKLKIQCLBGGIWEQSCIPVYCEPPPVFEGMYE 1607  
 DB 1381 VGTICKECKRPGYVVAESAGKVRNKLKIQCLBGGIWEQSCIPVYCEPPPVFEGMYE 1440  
 QY 1608 CTNGFSLDSQCVLNCNERERKPLICTKEGLMTQEPFLCENLOGECPPPSELSVVEYKC 1667  
 DB 1441 CTNGFSLDSQCVLNCNERERKPLICTKEGLMTQEPFLCENLOGECPPPSELSVVEYKC 1500  
 QY 1668 EOGYIGAVCSPLCVIPSPDPVMLPENITADTLEHNMPEYKVISVCTGRORHNPDPVLV 1727  
 DB 1501 EOGYIGAVCSPLCVIPSPDPVMLPENITADTLEHNMPEYKVISVCTGRORHNPDPVLV 1560  
 QY 1728 HCTIOSCEPFOAGNCDTINNRAYCHYDGGCCSSTLSSKKVIPPADCDLDECTCRDPKA 1787  
 DB 1561 HCTIOSCEPFOAGNCDTINNRAYCHYDGGCCSSTLSSKKVIPPADCDLDECTCRDPKA 1620  
 QY 1788 EENQ 1791  
 DB 1621 EENQ 1624

RESULT 4  
 O8BJG6 PRELIMINARY; PRT; 1214 AA.  
 ID O8BJG6  
 DT 01-MAR-2003 (Tremblrel. 23, Created)  
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)



DT 01-MAR-2003 (Tremblrel. 23, last annotation update)  
 DE Similar to pregnancy-associated plasma PREPROTEIN-A2  
 (Fragment).  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NC NCB1\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Dorsal root ganglion;  
 RX MEDLINE=22354683; PubMed=12466851;  
 RA The FANTOM Consortium,  
 RA The RIKEN Genome Exploration Research Group Phase I & II Team;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs."  
 RL Nature 420:563-573(2002).  
 DR EMBL; AK084006; BAC39093.1; -;  
 FT NON\_TER 1  
 FT NON\_TER 1214  
 SQ SEQUENCE 1214 AA; 134983 MW; AF4204160F05A285 CRC64;

Query Match 53.6%; Score 5281.5; DB 11; Length 1214;  
 Best Local Similarity 79.1%; Pred. No. 0;  
 Matches 960; Conservative 102; Mismatches 149; Indels 3; Gaps 2;

QY 164 KGSAMATTTLTTLTTLNEKPEPTEORRGMAKSRQRAADGGD-SGISHPQW 222  
 DB 1 EGSSEIAVSTIAFAVOKGPTATQKKGSKSLPQVQKQEDMTGDPQNTPOGQLM 60  
 QY 223 PKHSLHRYKSPSPESNONGSGSYREAFENSOVLPLTFSGRRERLLRPEVLAET 282  
 DB 61 PKPELKHGSDSPLEGTIQNGGASLRVETPNSOGGLPVLYFTGKRELLRPEVLAET 120  
 QY 283 PREAFTEAMVKEGQNNPAILAGVFDNCSHTVSDKMAIGRSGDKKRAARPPFSL 342  
 DB 121 PREAFTEAMVKEGQNNPAILAGVFDNCSHTVSDKMAIGRSGDKKRAARPPFSL 180  
 QY 343 CTDRVKKATTLISHSRYPQGTWTHVAATVYGRHMAIVDGTQVASSLDGSPINPMAS 402  
 DB 181 RTRDKMKATTLVTHGSKRYQPMWTHVAATVYGRHMAIVDGTQVASSLDGSPINPMAS 240  
 QY 403 CRSLILGDSSEDDGHYFRGHAGTLVFNSTALPQSHFQSSSHSGEATDLVLTASFEP 462  
 DB 241 CRSLILGDSSEDDGHYFRGHAGTLVFNSTALPQSHFQSSSHSGEATDLVLTASFEP 300  
 QY 463 VNTENVPFRDEKPRLEVLQGFPEPEPILSPLOPPLCGQTVCNVELISQYNGIWPILRG 522  
 DB 301 LMEQMAPFRDLYPRLEVLQDSSEPELILSPLOPPLCGQTVCNVELISQYNGIWPILRG 360  
 QY 523 KVIROYVNNICDEGLNPVSEIOIRLOHBALEAFRNYISMOQSHOVHNSTLRHRY 582  
 DB 361 KVIROYVNNICDEGLNPVSEIOIRLOHBALEAFRNYISMOQSHOVHNSTLRHRY 420  
 QY 583 LVNCEBSKIGNDHCDECEHPLTGYDGDGCRLOGRCYSMNRRLGLCHVECNMNLNFD 642  
 DB 421 LVNCEBSKIGNDHCDECEHPLTGYDGDGCRLOGRCYSMNRRLGLCHVECNMNLNFD 480  
 QY 643 DCCDPOVADYRKTCCFPDPSKRAYMYKELKEALQLNSTHFLNITYPASVREDLAGAAT 702  
 DB 481 DCCDPOVADYRKTCCFPDPSKRAYMYKELKEALQLNSTHFLNITYPASVREDLAGAAT 540  
 QY 703 PMDKAVTHLGGIVLSPAYYGMGHTDTMHEGVHVLGLVHFKVGSERSCNDPCKEY 762  
 DB 541 PMDKAVTHLGGIVLSPAYYGMGHTDTMHEGVHVLGLVHFKVGSERSCNDPCKEY 600  
 QY 763 PSNETDLCADTAPTPKSELCREPEPTSDTCGTRPAGAFITMYSTYDNDCTDNFTPNQ 822  
 DB 601 PSNETDLCADTAPTPKSELCREPEPTSDTCGTRPAGAFITMYSTYDNDCTDNFTPNQ 660  
 QY 823 VAMHCYLDLVYQOMTESKPT--PIPIPMVYGTQNKSLTIHMLPISGVYTDRAAGSL 880  
 DB 661 VAMHCYLDLVYQOMTESKPTPIPIPIPMVYGTQNKSLTIHMLPISGVYTDRAAGSL 720

QY 881 CGACTEDGTFROYVHTASSRRVCDSSGYWTPREAVGPDPVDGCEBSLQAMSEVHLHYM 940  
 DB 721 CSACTEDGTFROYVHTASSRRVCDSSGYWTPREAVGPDPVDGCEBSLQAMSEVHLHYM 780  
 QY 941 NMTVPCEPTGCSLELLFQHPVQADTLTLVNTSFPMESQVLPFTLELTKESVHLGPLD 1000  
 DB 781 NMTVPCEPTGCSLELLFQHPVQADTLTLVNTSFPMESQVLPFTLELTKESVHLGPLD 840  
 QY 1001 TPCDIPLTIKLVNDGVSQVKKYTPPERTEIDAALLTSQPSHLSGSGCPVRQVLRDPP 1060  
 DB 841 TPCDIPLTIKLVNDGVSQVKKYTPPERTEIDAALLTSQPSHLSGSGCPVRQVLRDPP 900  
 QY 1061 PASGLPVVVTTHSHRKTDEVTVPQWYQVQVLAAGSGEASPPPLNHIHGAPYCGDGKY 1120  
 DB 901 FSSGLPMVYTHHRKKTDEVTVPQWYQVQVLAAGSGEASPPPLNHIHGAPYCGDGKY 960  
 QY 1121 SERLGECDGDIVSGDCSKYCELEGFNCVGEPSLCYMEBGDGIPEPERKTSIVDCG 1180  
 DB 961 SERLGECDGDIVSGDCSKYCELEGFNCVGEPSLCYMEBGDGIPEPERKTSIVDCG 1020  
 QY 1181 IYTPKGYLDQMAVTRAYSHEDKKKCPVSLVTGEPHSLICTSYHPLPNHRPLTGMFPVVA 1240  
 DB 1021 IYTPKGYLDQMAVTRAYSHEDKKKCPVSLVTGEPHSLICTSYHPLPNHRPLTGMFPVVA 1080  
 QY 1241 SENETQDRSEOPESGLKKEDEVMLKVCNRPGEARAFIFLTTDGLVGEHQOPTYLY 1300  
 DB 1081 SUKQODASEQAKSDQKDNELWLEVCNRPGEARAFIFLTTDGLVGEHQOPTYLY 1140  
 QY 1301 LIDVRSNHSLSGTYGLSCQNPILINVTGHQVLFHHTSVLLNFSPPRGVISAVALRTS 1360  
 DB 1141 LIDVRSNHSLSGTYGLSCQNPILINVTGHQVLFHHTSVLLNFSPPRGVISAVALRTS 1200  
 QY 1361 SRIGASAPNCISE 1374  
 DB 1201 SRIGASAPNCISE 1214

## RESULT 5

Q96PH7 ID Q96PH7 PRELIMINARY; PRT; 826 AA.  
 AC Q96PH7;  
 DT 01-DEC-2001 (Tremblrel. 19, Created)  
 DT 01-DEC-2001 (Tremblrel. 19, last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, last annotation update)  
 DE Pregnancy-associated plasma protein E2.  
 GN PAPPE.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 NC NCB1\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21482666; PubMed=11597188;  
 RA Page N.M., Butlin D.J., Lomhaersong K., Lowy P.J.;  
 RT "The Characterization of Pregnancy Associated Plasma Protein-E and the  
 RT Identification of an Alternative Splice Variant."  
 RL Placenta 22:681-687(2001).  
 DR EMBL; AF342990; AAL17780.1; -;  
 DR InterPro; IPR000800; Notch.  
 DR SMART; SM00004; NL; 1.  
 DR PROSITE; PS00142; ZINC PROTEASE; 1  
 SQ SEQUENCE 826 AA; 92032 MW; FBFA09D652065D1 CRC64;

Query Match 44.6%; Score 4394; DB 4; Length 826;  
 Best Local Similarity 99.9%; Pred. No. 0;  
 Matches 808; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 MCKIIRISLIALAGALCSANSELGWTTRKSLVEREHLNOVLLEBERCYLAKXRRPRA 61  
 DB 1 MCKIIRISLIALAGALCSANSELGWTTRKSLVEREHLNOVLLEBERCYLAKXRRPRA 60  
 QY 62 SPQHHLFGVYPSRAGNYLRPVYGEQEIHHHTGSKSDTSGNANSLVPRDLETNPALRG 121

```

Db      61 SPQHHLFGVYPSRAAGNLRYPVGEQSHHTGSKPTBEGNANSLVPEPDLTENPAGLRGA 120
Qy      122 VEEPAAPWVDSPIGOSELLIGDDAYIGNORSKESLCEAGIQKGSAMAAATTTTALFTTLN 181
Db      121 VEEPAAPWVDSPIGOSELLIGDDAYIGNORSKESLCEAGIQKGSAMAAATTTTALFTTLN 180
Qy      182 EPPKPEORRMASROROROWKRAERAGQDSDGSSHPQWPHKSLKHKRKSPRESNQ 241
Db      181 EPPKPEORRMASROROROWKRAERAGQDSDGSSHPQWPHKSLKHKRKSPRESNQ 240
Qy      242 NGEGSYREAEETNSQVGLPIFYSGRRERLLPREVLAIEPREAFTVEAMVKEGGONN 301
Db      241 NGEGSYREAEETNSQVGLPIFYSGRRERLLPREVLAIEPREAFTVEAMVKEGGONN 300
Qy      302 PAIAGVFDNCSTHSYKGMALGIRSGDKYKADARFPFSLCTDRVKKATILSHSRYP 361
Db      301 PAIAGVFDNCSTHSYKGMALGIRSGDKYKADARFPFSLCTDRVKKATILSHSRYP 360
Qy      362 GWTWVAATYDGRMALVYDGTQVASSLDOSGPLNSPFMASCRSLLGDSSEDEHYFRG 421
Db      361 GWTWVAATYDGRMALVYDGTQVASSLDOSGPLNSPFMASCRSLLGDSSEDEHYFRG 420
Qy      422 HLGTLVFWSTALPQSHFOHSSOHSGEBAITDVLITASPEPVNTEWVPFDEKYPRLV 481
Db      421 HLGTLVFWSTALPQSHFOHSSOHSGEBAITDVLITASPEPVNTEWVPFDEKYPRLV 480
Qy      482 QGFEPEPEIISPLQPLCGQTCVNNELISQYNGYWPRLGKVIYQVNVICDEGLNPI 541
Db      481 QGFEPEPEIISPLQPLCGQTCVNNELISQYNGYWPRLGKVIYQVNVICDEGLNPI 540
Qy      542 VSEBOIRLOHEALNEAFSRNYSWQLSVHOVNSTLHRVVLVNCESPKIGNDHCEDE 601
Db      541 VSEBOIRLOHEALNEAFSRNYSWQLSVHOVNSTLHRVVLVNCESPKIGNDHCEDE 600
Qy      602 HPLTGYDGDGCRLOGRQCYSNMRDGLCHVECNMNLNFDGDCDPOVADVRKTCFDPDS 661
Db      601 HPLTGYDGDGCRLOGRQCYSNMRDGLCHVECNMNLNFDGDCDPOVADVRKTCFDPDS 660
Qy      662 PKRAYMSVKELKEALQJNSTHFLNIYFASSVREDLAAGATPMDKXAVTHLGGIVLSPAY 721
Db      661 PKRAYMSVKELKEALQJNSTHFLNIYFASSVREDLAAGATPMDKXAVTHLGGIVLSPAY 720
Qy      722 YGMGHDITMHEHGVILGVHFKVGSERESCNDPCKETVPSMETGDLCDTAPTPKSE 781
Db      721 YGMGHDITMHEHGVILGVHFKVGSERESCNDPCKETVPSMETGDLCDTAPTPKSE 780
Qy      782 LCREPEPTSDTCGTRFPGAFPTNMSYT 810
Db      781 LCREPEPTSDTCGTRFPGAFPTNMSYT 809

```

## RESULT 6

```

Q8R4K8 PRELIMINARY; PRT, 1545 AA.
ID Q8R4K8 AC Q8R4K8
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Pregnancy-associated plasma protein-A (Fragment).
GN PAPP-A.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., Thomsen A.R., Laursen L.S., Olsen I.M.,
RA Haaning J., Sottcutt-Jensen L., Haaning J., Giudice L.C., Conover C.A.,
RA Oxyig C.;
RT "Expression of recombinant murine PAPP-A and a novel variant (PAPP-A1)
RT with differential proteolytic activity.";
RL Eur. J. Biochem. 0:0-0(2002).

```

```

DR EMBL: AF439513; AAM12687.1; -.
DR MGD: MGI:97479; Pappa.
DR InterPro: IPR006558; Lamg_1ike.
DR InterPro: IPR000800; Notch.
DR InterPro: IPR000436; Sushi_SCR_CCP.
DR InterPro: IPR006025; Zn_MTPetidae.
DR Pfam: PF00084; Sushi; 5-.
DR SMART: SM00032; CCP; 4-.
DR SMART: SM00560; LamGL; 1.
DR SMART: SM00004; Nt; 3.
DR PROSITE: PS00142; ZINC_PROTEASE; 1.
FT NON TER 1
SQ SEQUENCE 1545 AA; 172584 MM; B820BF9998245419 CRC64;
Query Match 40.3%; Score 3974; DB 11; Length 1545;
Best Local Similarity 47.3%; Pred. No. 0;
Matches 733; Conservative 280; Mismatches 486; Indels 50; Gaps 24;
Qy 263 LYFSGRRERLLPREVLAIEPREAFTVEAMVKEGGONNPAIAGVFDNCSTHSYKGMAL 322
Db 15 LYFSGRRERLLPREVLAIEPREAFTVEAMVKEGGONNPAIAGVFDNCSTHSYKGMAL 72
Qy 323 LGIRSGDKYKADARFPFSLCTDRVKKATILSHSRYPQCTWTHVAATYDGRMALVYD 382
Db 73 MGJHTTSDQGNRDPFRFFSLKTDRAKVTITIDHRSYLRQWVHLAATYDGRMLKLYMG 132
Qy 383 TOVASSLDOSGPLNSPFMASCRSLLGDSSEDEHYFRGLTGLVFWSTALPQSHFOHSS 442
Db 133 AOVATSAEOVGIFSPITOKCKVLMIG--SALNHNFRGTHIFSLMKVARTREIVSDM 190
Qy 443 QHSSGEBAITDVLITASPEPVNTEWVPFDEKYPRLV--LQFEPEPEIISPLQPLCG 500
Db 191 ETRGHLTPRLQQLQGNMVDVVKRTWSBMKDGNSPOVEFSNAHGFLLD---TTLPEPLCG 246
Qy 501 QTVCDNVELISQYNGYWPRLGKVIYQVNVICDEGLNPIVSEBOIRLOHEALNEAFSR 560
Db 247 QTLCDNTEVSISSNQLPSFRQPKVVRVYVNIYDHDHNEPTVSWQOIDFHOOLAEAFQ 306
Qy 561 YNISWQLSVHOVNSTLHRVVLVNCESPKIGNDHCEDEHPLTGYDGDGCR--LQGRQCY 619
Db 307 YNISWELVAVNINSSIRNHLITLNCDSIKIGKCDPBCNHTLTHGDDGDCQOLRPAF 366
Qy 620 SNMRDGLCHVECNMNLNFDGDCDPOVADVRKTCFDPDSPKRAYMSVKELKEALQJLN 679
Db 367 MKQQNGVCMDNCNRYEFNFDGEGCDPDLTDVTKCFDPDSHRAVLQVNLKILRLD 426
Qy 680 STHEFLNIYFASSVREDLAAGATPMDKXAVTHLGGIVLSPAYYGMGHDITMHEHGVIL 739
Db 427 GSTHLNIFPANSSEELAGVATPMDKELMLHGGIVLNPSPYGIPIGHTHTMHEIGHSL 486
Qy 740 GLVHFKVGSERESCNDPCKETVPSMETGDLCDTAPTPKSELCREPEPTSDTCGTRFP 799
Db 487 GLVHIFRIGISELSCSDPCMETEPSEFETGDLCDTNPAPRHKFCGDPGPNDDTCGFHGF 546
Qy 800 GAPFTNMYTDCNCTDNFTPNQVAMHLYLDVYQWTESSRKPTPIPIPNVIGQTNKS 859
Db 547 NTPYNNMSTADDDCTDSTFTPNQVSRHLCYLDLVYQWQSRKRPAPALAPQVGHMTMS 606
Qy 860 LTIHMLPPIGCVYDRAAGSLCGACTBEDGTFRQYVTHASSRRVCDSSGYTTPBEAVCPD 919
Db 607 VMLEWFPPIIDGHFFERELGSAACDLCEGRILVOYAFNASSPMPCGSGPBGHMSPREABEHPD 666
Qy 920 VDQPCESLQAMPREVNLHMMNTVPCP--TEGCSLELLPQHPQADTLTLMVT--SFFME 976
Db 667 VEQPCKSSTVWSSNANVHNPVAPCPBEQGCLELEFRRYPLPESLTIWVTVSSDMD 726
Qy 977 SSQVLFTEILLENKESVHLGLPDTCDIPLLTIKLN--VDKGVSGVKKYTFDEBIEIDAL 1035
Db 727 SSGAVNDIKILLTISGKNSISGPNQVFCDIPILIRLRVGEVYGIQYITIDHNHEIDAM 786
Qy 1036 LITQSPHPLSCGCPVRYQVLRDPFASGLPVVVTSHRKFTDEVTPGQMYQYQVLAEA 1095
Db 787 LITFVDSPLCLQCKPLQKYLRLDPPLEDAVSL--HINRRFMDTLKLGSGVYQVRIITIS 845

```

QY 1096 GSELGASPRYLNIHGAPYCGDGKVSRLGECDDGLVSGDGSKYCELEEGFNCGEP 1155  
 DB 846 GNESESPALAYTHGSGYCGDGYIQKQGEBCDDMKVNGDGLSCXQEVSNCLDER 905  
 QY 1156 SLCTYMEBGDICEPFEKTSIVDCGITYPGYLDQMATRAYSSHEDKKCPVSLVTGER- 1214  
 DB 906 SRCYFHGDGCEFEKTSIKQCGVYTPQGFIDQMSASVSHQD-QQCPGVVIGQPA 964  
 QY 1215 HSLICTSYHDPILPHNRPLTGMFPCVASENPTODRSQRSGSLKKEDEVLKXCFENRGE 1274  
 DB 965 ASQVCKTKYIDLSEGISQHAMPTCTT-----YPTVHL-PQTFWLDQTYFSQPMV 1013  
 QY 1275 ARAIFELTDLGVPEHQOPTYVLTLDVKGSNHSLGTYGLSCOHNPILINVTHQNVL 1334  
 DB 1014 AAIVILHVTGDTGYGQKQETISVQLDTRKQSHDGLHVLSCRNPLIIPVHDSQP 1073  
 QY 1335 FHHITSLVLFSSPRVIGISAVALTSTRIGLSAPNSCISEDEGQNHQSGCIRPCGKOD 1394  
 DB 1074 FYHSAVHVSFSSPLVAISGVALRSFDPNFPVTLSSC-QRGETYSRAEQSCVHFACQAD 1132  
 QY 1395 SCPSLLDHDADVNCST-----IGPGLMKCAITGOEGALQASSGQYIRPMQ--KEILLTC 1448  
 DB 1133 -CEELAVGNBS-LNCSNNHHYHG--AQCTVSCQTYVLIQIQDDELKISQVGPSTIVTC 1187  
 QY 1449 SSGHWDQNVSLPVDGCVDPPLVNYANFSCSEGTKEFKRCSISCVPRAKLOGSLPWLTC 1508  
 DB 1188 TEKMKKQVACFVYDQCI RPHHHVYASFSCEGTTGRCRCSFOGHRPADLKNNSLTLC 1247  
 QY 1509 LEBGLMSLPVYCKLBEDARPIILANLLPHCLQDNHDTVGTICKYCKPGYVVASAEG 1568  
 DB 1248 MEGSLSPFALCELMCLAPRPVADLQTPARCEENHKGVSFCYKCKPGYVPGSSR- 1306  
 QY 1569 KVNKLKLTQCEGGTMEGSSCI PVNCEPRPVPEGEYECTNGFSLDQCVLNC-----N 1623  
 DB 1307 KSKRRAPKQCTQDGSQWQETCVPVTCDDPPPKFHGLQCTNGQFNSERIKCEDSDAS 1366  
 QY 1624 QEREKPLILCTKEGLTQEFKLCENLOGECPPESELS--VEYKCEQGYIGAVCSPLCY 1682  
 DB 1367 QGHSNIIHCRKQKGTSGSFHVCREMGGC-SAPNOLNSMLKQICPBGVATISBICALSL 1425  
 QY 1683 IPPSDVPLPENITADTLEHMEBVKVQSIYCTGRQRHDPVYLHVICQSEFPADGWC 1742  
 DB 1426 DHNSESLIFVNLVTRVDI RPHMNPTRQRIYCTAGLQWPHPALIHCVKCEPFMGDNVC 1485  
 QY 1743 DTNNRAVCHYDGDCCSSSTLSKKVIRPADCDL-DECTCRPKAEEN 1790  
 DB 1486 DAINNRAFCNYDGDCCCTSTVTKVTPFPMSCDLQNDACRDEAGEH 1534

RESULT 7  
 Q8R4K7 PRELIMINARY; PRT; 1574 AA.  
 AC Q8R4K7  
 DT 01-JUN-2002 (TREMBLrel. 21, Created)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE Pregnancy-associated plasma protein-A variant (Fragment).  
 GN PAPPA.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Soe R., Overgaard M.T., Thomsen A.R., Laursen L.S., Olsen I.M.,  
 RA Haaning J., Sottrup-Jensen L., Haaning J., Gludice L.C., Conover C.A.,  
 RA Ovig C.;  
 RT "Expression of recombinant murine PAPPA and a novel variant (PAPPA-I)  
 RT with differential proteolytic activity";  
 RL Eur. J. Biochem. 0:0-0(2002).  
 DR EMBL; AF439514; AAM12688.1; -.  
 DR MGD; MGI:97479; Pappa.

DR InterPro; IPR006558; LamG\_1ike.  
 DR InterPro; IPR000800; Notch.  
 DR InterPro; IPR000436; Sushi\_SCR\_CCP.  
 DR InterPro; IPR006025; Zn\_MTPeptide.  
 DR Pfam; PF00084; sushi\_5.  
 DR SMART; SM00032; CCP; 4.  
 DR SMART; SM00560; LamGL; 1.  
 DR SMART; SM00004; NU; 3.  
 DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 FT NON TER 1 1\_PROTEASE; 1.  
 SQ SEQUENCE 1574 AA; 176105 MW; AE53F0187E6EB7B CRC64;  
 Query Match 40.1%; Score 3949.5; DB 11; Length 1574;  
 Best Local Similarity 46.5%; Pred. No. 0;  
 Matches 733; Conservative 280; Mismatches 486; Indels 79; Gaps 25;  
 QY 263 LYSGRERLLRPEVLAETPREAFTVEAVKVEGQNNPAILAGVFNDCSHVSDKGA 322  
 DB 15 LYPFGGEQLRLRADL-ELPRDAFTLQVWLRAGGQKSPAVITGLYDKCSYTSRDRGW 72  
 QY 323 LGRSGKDKGRDARREFSLCTDRVKKATLLSHRYOQGTWTHVATYDGRMALVYDG 382  
 DB 73 MGIHTTSDQGNRPRIYFSLKTRARKVTTIDHRSYLPQWVHLATATYDGRMLKLYMG 132  
 QY 383 TVYASSLDQSGPLNSPFMASCRSLDGSSEDEGHYFRGHAGTLVFWSTALPOSHFQHS 442  
 DB 133 AQVATAEQVGGIFSLTKCKVLMIG--SALNHFRRGHIEHPSLMKYARTQREIVSDM 130  
 QY 443 QHSSGEEATDLVLTAFEPVNTWVPRDEKYPRLV--LOGFEPEELISPLQPLCG 500  
 DB 191 ETRGLHTPLRQLLQENMNVNKRKTSPMKDGNSPQVEFSNAHQFLD---TWLBPPLCG 246  
 QY 501 QTVCDNVELISQVNGWPLRGEVIRYQVYVNTCDDGLAPYVEEQIRLOHEALNFRSR 560  
 DB 247 QTLCDNTEVISTNQVPSFRQPVVRYRVVNIYDHHNEPTVSWQIDFQHQDLAEFQH 306  
 QY 561 YNISQLSYHQVNSTLRHRYVNLNCEPSKIGNDHCDPECEHPLTYGDDGDCR-LOGRCY 619  
 DB 307 YNISWELEVLNINSSLRRLILANDISKIGBEKCDPECNHPLTGHDDGDCQLRIPAF 366  
 QY 620 SNMRDGLCHVECNMNLNPDGDCDPOVADVRYKTCFDPDSP----- 662  
 DB 367 MKKQNGVCDMDNRYRNFEDGECDDPITDVTYKTCFDPDSBQRSIRKRAVVEESWL 426  
 QY 663 -----KRAMSYKELKEALQLNSTHNLNYPASSVREDLAAGATWPDADAVT 710  
 DB 427 PHGQAKRKRKTRAVLDVNLKNILRLDGSTHLNIFPANSSEELAGVATWPDKEALM 486  
 QY 711 HLGIVLSPAYYGMPTHTMIHEVGHVGLYHVFQGVSEBESGNDPCKETVPSMETGL 770  
 DB 487 HLGIVLNSPFGIIPHTHTMIHEIGSLGLYHIFGISEIGSCSPCMETETSPETGDL 546  
 QY 771 CADTAPTPKSELCREDEPTSDTCGFTTRPGAPPTNMYSTDNCTNFTPNQVARNCYL 830  
 DB 547 CNDTNAPRKHKFCGDDGPGNDTCGFHGFNTFPYNNMYSVADDDCTDSFPNQVSRMHCYL 606  
 QY 831 DLVYQQWTSRKPTRPIPIPMVIGQNKSLTIHMLPPIGIVVDRAAGSLGACTEDGT 890  
 DB 607 DLVYQSWQSPRRKAPALAPQVVGHTMDSVLEWPFPIGHFEREISGACDLCLEGRIL 666  
 QY 891 ROYVHTASRRVCDSSGYTPEEAVGPVPVDDPCBSLQAWSPEVLYHNNMVTVPCTE 949  
 DB 667 VOYAFNASSPBMCGPBGHMSPRABEHPVDVDCSKSVTWSNSAVNPHTVPPACDEPQ 726  
 QY 950 GSELLELFQHPVOADTLTLMTV--SFFMESSQVLPFTELLLENKSVHLGAPLDFCDPL 1007  
 DB 727 GCVLELFRYPLVPESLTIWTFVSSDMDSGAVNDIKLLTISGKNISLQPGVPCDPL 786  
 QY 1008 TIKLH-VDDKVGAVKYUTDEREIDALLTQPHRPLSCGCPVAVYQLDRPPRSGP 1066  
 DB 787 TIRLRVGEVYGLIYTLDEHLEIDAAMLTVSDSPLLCQCKPPLQYKVLDRPPLLEDA 846  
 QY 1067 VVHTSHRFTVEVTPQMYQYVLAEGSELGASPRYLNIHGAPYCGDGKVSRLG 1126

```

Db      847 SLI-HLNRFFMDTLKLGSGVYQYRIITISGNESESPSAALYTHGSGCGVGIQXQGE 905
Qy      1127 ECDGDGLVSGDGCSCVCELEEGNCVGPBSLCVNYBEGGICEPPERTSTYDCIYPRKG 1186
Db      906 ECDMDKRVNGGCGSLFCQKQEVFNCDIPSSCYHDGDMKEEFQKTSIDCCVYPRQG 965
Qy      1187 YLDQMATRAVSSHEDKKCPVSLVTGEP-HSLICTSYHPDLPNHRPLTGWPCVASENET 1245
Db      966 FLDDQMANASVSHD-QCCPGWVYIGDPAQVCRKVIDSEGISQHAMVPCITIT----- 1020
Qy      1246 QDRSEBPESGLKKEDEVMKVCNRRPEEARITPLTTDGLVGEHQPTVTLITDVR 1305
Db      1021 -----YPRYHL-QPTTFMLQTFYSGPWWAAVAIIVHVTDTGYGDQOETISVQLDITK 1073
Qy      1306 GSNHSLGTYSLSGCHNPILIVTHNQNLFHHITTSVLLNSPVGISAVLPTSSRIQL 1365
Db      1074 DQSHDLGHLVLSGNNPLIPVHDLQPFYHQAQVHVSFSSPLVALSGVALRSFDPNDP 1133
Qy      1366 SAPENCSISEDEGQNHQGCSCIRPCQKQDCPSLLDHDADVNCSTSGPGLMKCAI 1421
Db      1134 VTLSSC-QRGETVPABGSCVHFACQAD-CEPLAVGNAS-LNCSNNHNG---AQCTV 1187
Qy      1422 TCGRGFALQASSGQYIRPMQ--KEILITGSSGHWQDVNSCLPVDGVPDPSLVANFSC 1479
Db      1188 SCOTGVYQLQIRDELKISQVGPSTIVCTEGKNNKQVACBPVDCGIPDHNVVAASFS 1247
Qy      1480 SEGKPLKRGISCVPRKLGSLPWLCTLEFDGLMSLPVYVCKLECAPIILIANLLP 1539
Db      1248 PEGTTFGRGCFQRHRAQLGNNSPFLTCMEDGLMSPEALCELMCLAPVPVPAADQTA 1307
Qy      1540 HCLDNDHVGITCYECKEPGYVAESABGRNKLKIQCLEGGIWEQSGCIPVCEPP 1599
Db      1308 RCERNKHKVGSFCYKCKKRGVHVGSSR-KSKKRAFKQCTQDQSGWQGTCTVPTCDPP 1366
Qy      1600 PVEGMECTNGFSLDSQCLNCG---NQEREKLPILCTEGLMTQEFKLCENIQEBCP 1654
Db      1367 PKFGLYQCTNGFQFNSCEKICKEDSDASQGRGSIHCRKQGTWGSFHVCREMQGC- 1425
Qy      1655 PRPELINS-VEYKQEQGIGAVCSPLCVIPPSDPVMLPENITADLEHMMPEPKVQSI 1713
Db      1426 SAPQLNSNLKLCQCPDGAISECAISCLDHNSSIILIPVLYTRADIPHMNPTFRVQIV 1485
Qy      1714 CTGRROWHPDVLVHICQSCPEFQADGWCPTIRNNRAYCHYGGDCSSSTLSKKVIIPAA 1773
Db      1486 CTALQWYPRHALHCVKGCPEFMGDNYCDAINNRAFCNYGCGDCTSTVTKKVTTPPM 1545
Qy      1774 DCDI-DECTCRPKAEN 1790
Db      1546 SCDLQNDCAACRDPQAQEH 1563

RESULT 8
Q9ES06 PRELIMINARY; PRT; 1367 AA.
ID Q9ES06 AC Q9ES06;
DT 01-MAR-2001 (TREMblrel. 16, Created)
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
DE Pregnancy-associated plasma protein-A (Fragment) .
GN PAPP.
OS Mus musculus (Mouse) .
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RA Hourvitz A., Kuwahara A., Hennebold J.D., Negishi H., Tanaka M.,
RA Wigger A., Erickson G.F., Adashi E.Y.;
RT "The regulated expression of insulin-like growth factor binding
RT protein-4-endopeptidase (Pregnancy-associated plasma protein-A) in the
RT rodent ovary: A marker of the dominant follicle and of the corpus

```

```

RT      luteum."
RL      Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR      EMBL: AF258461; AAC09799.1;
DR      MED: M46.001;
DR      MED; MGI:97479; Pappa.
DR      InterPro; IPR000800; Notch.
DR      InterPro; IPR000436; Sushi_SCR_CCP.
DR      InterPro; IPR006025; Zn_MTPetidae.
DR      Pfam; PF00084; sushi; 5.
DR      SMART; SM00032; CCP; 4.
DR      SMART; SM00004; NL; 3.
DR      PROSITE; PS00142; ZINC_PROTEASE; 1.
FT      NON TER 1
SQ      SEQUENCE 1367 AA; 152644 MW; A46BC5B0449042A5 CRC64;

Query Match      35.7%; Score 3517; DB 11; Length 1367;
Best Local Similarity 47.1%; Pred. No. 2.7e-283;
Matches 650; Conservative 248; Mismatches 419; Indels 64; Gaps 23;

Qy      431 TALPQSHFQSSQHSQSGSEATDLVLASFEPVNTVPPRDEKYPRLV--LQGFEP 488
Db      19 TPLPQ-----LILQENWDNKRKTSYMKDQNSQVFEESNAHGFLLD- 59
Qy      489 ELISPLQPLCGQTVCDNVELISQYNGWPLRGEKVIRYQVNI CDDEGLNPVSEQIR 548
Db      60 ---TNLEPPLCGQTLCDNTEVISYNQLPSFRQPKVRYRVNVIYDDHNPVTSWQID 116
Qy      549 LQHEALNEASRYNISQVLSVHOVHNSLTRRVLVNVCESKLGNDCECEHPLTGYD 608
Db      117 FOHQQLLEAHOYHIVISLEVLINSSILHRLLANCDISKIDCEKCDCECHTLTGHD 176
Qy      609 GQDGR-LQGRCYSNRRDGLCHVENNMNLNDFDQDCDQVADVRKTCFDPSPKRAYM 667
Db      177 GQDGRQLRYAFMKKQNGVCMDCNVERNPFQGGECDDIDYVTKCTCDPSPHAYL 236
Qy      668 SVKELKALQNLSTHFINITYPASSVREDLAGAATVPWDKQAVTHLGIVLSPAYYGM 727
Db      237 DVNELKQILRLDGSHTLINFANSSBEELAGVATVPWDKALMHLGIVLNPFGYGI 296
Qy      728 TDMTHHGVHVLGLYHFKVRSERESGNDPCKETVPSEMGDI CADPAPTPKSELCRE 787
Db      297 THTHIHIGSLGLYHIFRQISQSCDPCMETEPSEFEGDLCNDTPAPKHKFCGDP 356
Qy      788 PTPDTCGTFPFGAPFFNYMSYTDNCTDNFTPNQVARMHCYDLVLYQQTESRKPPIP 847
Db      357 PGNDTCGFHGFNTPNYNNFSYADDDCTDFTTRQVSRMHCYDLVLYQSQPBRKRPVA 416
Qy      848 IPRWVIGQTKSLTIHMLPPISGVYDRAAGSLCGACTEDGTRQVYHTASRRVCDSSG 907
Db      417 LAPQVGHITMDSVMLEWFPPIIDGHFFERELSGACDLCEGRILVQVAFNASSPMP 476
Qy      908 YMTPEEAVGPVDVQCEPSELQAMSPVHLVYHMMVYPCP-TGCGSLELLFQHPVQAD 966
Db      477 HWSPREAEGHPDVEQPKSSVRTWSPNSAVNPHTVPACPEQCYLEFRFPVLESL 536
Qy      967 TLWVT--SPFMESQVLPTEILLENKESVHLAGPLDFCDIPILTKLH-VDGKSGVKV 1023
Db      537 TIWVTPSSDMSDSSGAVNDIKLITSGKNTSISQPNVFCIPLITLRADGEBVYGIQY 596
Qy      1024 TFDERIBDALITSOHSPILCSGCRPVRYQVLRDPFASGLPVVYTHSRKFTDVEVP 1083
Db      597 TLDEHLIDAMLITSTVDSPLCLQCKPQYKVLADPRLDVLASLT-HLNRFFMDTL 655
Qy      1084 GQWYQVYVLAAGELGAEAPPLNHTGARYCGDGKXSEKLGECDDGDLVSGGCGKVC 1143
Db      656 GSVYQYRIITISGNESESPSAALYTHGSGCGVGIQXQGECDMDKRVNGGCGSLFC 715
Qy      1144 ELEEGFQVGPBSLCVNYBEGGICEPPERTSTYDCIYTPKGLDQMATRAVSSHEDK 1203
Db      716 KQEVSPNCIDPSSCYHDGDMKEEFQKTSIDCCVYTPQGFLDDMANASVSHD-Q 774
Qy      1204 KCPVSLVTGEP-HSLICTSYHPDLPNHRPLTGWPCVASENETQDRSEBPESGLKKEDE 1262

```

```

Db 775 QCEGWWVIGOPASQVCRKVIDLSEGISQHAMYPCTIT-----YFYHL-PQT 823
Qy 1263 VMLKVCNRRGEBAALIFILTLTGLVGEHQPFVTLVLTVDVSGNSHGLTYGLSCQHP 1322
Db 824 FMIQTTFSPQVMAAVALIHLVDTGTYYGQKQETISVQLDPTKQSDHGLHVLSCNNP 883
Qy 1323 LLIINVTQHONVLFPHHTTSVLLNFSSPRVGI SAVALARTSRIGLSAPANCISEDEQNHQ 1382
Db 884 LIIIPVVDLSQPFHQAVHVSFSSPLVALISGVALRFDPNPDVTLSSC-QRGTTYS PAE 942
Qy 1383 QSCIHRCGKODSCPSLLDHADVNTS----IGELMKCATTCQGFALQASSGOYIR 1438
Db 943 QSCVHFCAQAD-CPELAVGNAS-LNCSNMHYHG---AQTWVSCQGYVLQIQRDDLI 997
Qy 1439 PMQ--KEILLTSCSGHMDQVNSCLPVQGVDPDELVNYANPSSCEGKFLKRCISCVPP 1496
Db 998 KSOVGPSTIVCTGKNNKNOVACBPVCGIPDHNVVAASCSCEGTFPGRCFSFQGRHP 1057
Qy 1497 AKLQGLSPMLTCLLEDGLMSLPENVCKLECDAPPIILANLLPHCLQDNHVGITCKYEC 1556
Db 1058 AQLKGNNSPILTCMBEDGLMSFPEALCELMCLAPPVPAADLOTACREMKHVSFCYKIC 1117
Qy 1557 KPGYVAESAAGKVRNKLKIQLEGGIWEQSGCIPVCEPPPEPVEGMECTNGFSLDS 1616
Db 1118 KPGYHVGSSR-KSKKRAFKTQCTQDGSWQEGTCVPTCDPBPXKFGILYQCTNGFQNS 1176
Qy 1617 QCVLNC-----NOREKLPICTEGMLTQEFKLCENLOGECPPEPLNS-VYKCEOG 1670
Db 1177 ECRKCKDSASQGRGSIHCRDGTWGSFVHCRMOQGC-SAPQLNSMLLQCPDG 1235
Qy 1671 YGIGAVCSPLCVIPSPDPMLENITADTLEHMEPVKOSIVCTGRQMHDPVLVHCI 1730
Db 1236 YALGSECAISCLDINSSEIILPVVLYTRDIPHMNPFRVGRIVCTAGLQVPHALHCV 1295
Qy 1731 QSECPQADGWCDTINNRAYCHYDGSCCSSSTSSKKVIPPAAODI-DECTCRPAEE 1789
Db 1296 KGEPEFMDVYCDAINNRACFNVDGDCSTVTKTKVTFPMSCDILQNCACRDPAGE 1355
Qy 1790 N 1790
Db 1356 H 1356

```

RESULT 9  
Q9NUP4 PRELIMINARY; PRT; 396 AA.  
AC Q9NUP4; 01-OCT-2000 (TREMBlrel. 15, Created).  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE D65218.1 (Similar to PAPA (Pregnancy-associated plasma protein A))  
DE (Fragment).  
GN D65218.1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Donnelly S.;  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL031174; CAB2297.1; -  
DR MEROPS; M46.002; -  
DR InterPro; IPR000800; Notch.  
DR SMART; SM00004; NL; 1.  
FT NON TER 1  
SQ SEQUENCE 396 AA; 44619 MW; CD8CLBF38097506 CRC64;

Query Match 22.0%; Score 2173; DB 4; Length 396;  
Best Local Similarity 100.0%; Pred. No. 3.6e-172; Indels 0; Gaps 0;  
Matches 396; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
269 RERLLRPEVLAEIPREAFTEAVWVKEGQNNPAILIAGVFDNCSTHVSXGMLGIRSG 328

```

Db 1 RERLLRPEVLAEIPREAFTEAVWVKEGQNNPAILIAGVFDNCSTHVSXGMLGIRSG 60
Qy 329 KDKGKXDAARFESLCTDRYKATILISHSRVQGTTHVAATDGRHMLYVGTQVASS 368
Db 61 KDKGKXDAARFESLCTDRYKATILISHSRVQGTTHVAATDGRHMLYVGTQVASS 120
Qy 389 LDOSGPLNSPFMAKCSLLIGDSSDGHYFRGLGLTVWSTALPQSHFQSHSSGCE 448
Db 121 LDOSGPLNSPFMAKCSLLIGDSSDGHYFRGLGLTVWSTALPQSHFQSHSSGCE 180
Qy 449 BEATDVLTAEPFVNTWVPPRDEKYPRLVLOGEPEBEILSPIPLCGQTVCDNV 508
Db 181 BEATDVLTAEPFVNTWVPPRDEKYPRLVLOGEPEBEILSPIPLCGQTVCDNV 240
Qy 509 LISQVNGYPLREKRYIRQVYVNICDEGLNPIVSEBOIRLOHEALNEARSYNIWQLS 568
Db 241 LISQVNGYPLREKRYIRQVYVNICDEGLNPIVSEBOIRLOHEALNEARSYNIWQLS 300
Qy 569 VHQVHNSLTHRRVVLVNCSEPSKIGNDCPECEHPLTGYDGPDRLOGRCSYNNRRDGLC 628
Db 301 VHQVHNSLTHRRVVLVNCSEPSKIGNDCPECEHPLTGYDGPDRLOGRCSYNNRRDGLC 360
Qy 629 HVECNMMLNDFDDGDCDPPQVADVRKTCFDPDSPKR 664
Db 361 HVECNMMLNDFDDGDCDPPQVADVRKTCFDPDSPKR 396

```

RESULT 10  
O75997 PRELIMINARY; PRT; 219 AA.  
AC O75997; 01-NOV-1998 (TREMBlrel. 08, Created).  
DT 01-NOV-1998 (TREMBlrel. 09, Last sequence update)  
DT 01-JAN-2001 (TREMBlrel. 17, Last annotation update)  
DE D1774124.1 (Similar to pregnancy-associated plasma protein A  
DE (precursor).  
GN D1774124.1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Grafham D.;  
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL031290; CAA20358.1; -  
DR InterPro; IPR000800; Notch.  
DR InterPro; IPR000436; Sush1\_SCR\_CCP.  
DR SMART; SM00032; CCP; 1.  
DR SMART; SM00004; NL; 1.  
FT NON TER 1  
SQ SEQUENCE 219 AA; 24366 MW; 65CS311409C759DA CRC64;

Query Match 12.7%; Score 1255; DB 4; Length 219;  
Best Local Similarity 100.0%; Pred. No. 4.3e-96;  
Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
1573 KLIKIQLEGGIWEQSGCIPVCEPPPEPVEGMECTNGFSLDSQVLCNQRERKLPIL 1632  
Db 1 KLIKIQLEGGIWEQSGCIPVCEPPPEPVEGMECTNGFSLDSQVLCNQRERKLPIL 60  
Qy 1633 CTKEGLMTQEFKLCENLOGECPPEPPEPLNSVEKCEGYSIGAVCSPLCVIPSPDPMPL 1692  
Db 61 CTKEGLMTQEFKLCENLOGECPPEPPEPLNSVEKCEGYSIGAVCSPLCVIPSPDPMPL 120  
Qy 1693 ENTADTLEHMEPVKOSIVCTGRQMHDPVLVHCIQSECPQADGWCDTINNRAYCH 1752  
Db 121 ENTADTLEHMEPVKOSIVCTGRQMHDPVLVHCIQSECPQADGWCDTINNRAYCH 180  
Qy 1753 YDGGDCSSSTLSSKKVIPPAAODI-DECTCRPAEENQ 1791  
Db 181 YDGGDCSSSTLSSKKVIPPAAODI-DECTCRPAEENQ 219

```

RESULT 11
ID 09UK57 PRELIMINARY; PRT; 468 AA.
AC 09UK57;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Pregnancy-associated plasma protein A (Fragment).
GN PAPA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Embryo;
RA Olesen C., Hansen C., Hayashizaki Y., Bykov A., Tommerup N.;
RT "Partial sequence of Mus musculus Mus musculus pregnancy-associated
RT plasma protein A (Papa).";
RL Submitted (AFR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF260433; AAF70319.1; -.
DR MEROPS; M46.001; -.
DR MCD; MGI:97479; Pappa.
DR InterPro; IPR000800; Notch.
DR Pfam; PF00084; suah1; 5.
DR SMART; SM00032; CCP; 4.
DR SMART; SM00004; NL; 1.
FT NON_TER 1
SQ SEQUENCE 468 AA; 51134 MW; 85768B10D7D34FDA CRC64;

Query Match 11.2%; Score 1102.5; DB 11; Length 468;
Best Local Similarity 43.5%; Pred. No. 7.6e-83;
Matches 202; Conservative 85; Mismatches 156; Indels 21; Gaps 11;

QY 1340 SVLNFSPRGVIGAVLRTSSRIGLSAPNSCISEDEQNHQSCIRHPCGKQDCPSLT 1399
DB 2 AVHSFSPSPVLAISGVALRSFDFNPDTLSSC-QRGFTVSPABSCVHFAQCAAD-CDEL 59
QY 1400 LLDHADVNVCTST---IGPGLMKCAITCORGPALOASSGQYIRPMQ--KEILLTSSGHW 1453
DB 60 AVGNAS-LNGSNHNG---AQCTVSCQGTGYVLQIRGDBLKSQVGPSTITVCTSGKW 115
QY 1454 DQNVSCIPVDCGVDPRLVNTANFSCSEGTFLKRCISLCVPRAKLQGLSWLTCTEDGL 1513
DB 116 NKQVACEPVDCGIDPHHHVYAGSFSCPEGITFGRRCSPQCRHPRALQKGNNSFLTCEMDGL 175
QY 1514 WSLPEVVCLECDAPRIILNNLLPHCLODNHVGTICKECKPGYVVAESAGKVRNK 1573
DB 176 WSPFALCELMCLAPRPVPRNADLOTARCRKNHKGSCCKYCKRGINHPSSR-KSKKR 234
QY 1574 LKTIQCLEGIMBQSCIPVVCCEPRPVFEGMVECTNGFSLDSQVLNC-----NOEREK 1628
DB 235 AFKIQCTQDQSGWQGTGCVPTCDPRPKFHGLYCTNGFQNSERIKCEDSDASQKGS 294
QY 1629 LPICTKEGLTQEFKLCENIQSCPPPELNG-VEYKEQYIGIGAVSCPPLVTPSPD 1667
DB 295 NIIHCRKDGTSGFHYCREWQGC-SAPNQLNSNLKIQCDGVAIGSECAISCLDHNRG 353
QY 1688 PVMLPENITADTLEHMEPVVQSVIVCTGRQWHPDPVLVHCISCEPFQADGMDPTNN 1747
DB 354 SIILPVKLVKVDIDHMTVPRLGRICTAGLQMTVPPRLHICVAGCGPFMDNYCDALNN 413
QY 1748 RAYCHYDGGDCSSSTLSSKVIIPPAADCL-DECTCRDPKAEN 1790
DB 414 RAFCNVGGDCCTSTVTKKVTTPPMSCDLDQDCACRPPGAQH 457

```

```

RESULT 12
ID 08K423 PRELIMINARY; PRT; 354 AA.

```

```

AC 08K423;
DT 01-OCT-2002 (Tremblrel. 22, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Pregnancy-associated plasma protein A (Fragment).
GN PAPA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SvJ;
RA Qin X., Sexton C., Byun D., Strong D.D., Baylink D.J., Mohan S.;
RT "Differential Regulation of Pregnancy Associated Plasma Protein
RT (PAP)-A during Pregnancy in Human and Mouse.";
RL Growth Horm. IGF Res. 0:0-0(2002).
DR EMBL; AF510317; AAM44048.1; -.
DR InterPro; IPR006558; LamG_1like.
DR InterPro; IPR000800; Notch.
DR SMART; SM00560; LamGL; 1.
DR SMART; SM00004; NL; 2.
FT NON_TER 1
SQ SEQUENCE 354 AA; 40550 MW; 06093321EFB32805 CRC64;

Query Match 9.1%; Score 896.5; DB 11; Length 354;
Best Local Similarity 47.1%; Pred. No. 7.1e-66;
Matches 169; Conservative 59; Mismatches 122; Indels 9; Gaps 4;

QY 307 GVPDNCSHTVSDKMGALGIRSGDKGRDAPFFSLCTDRKATILISHRYQPGTWT 366
DB 1 GLYDKCSYTRSDRGWVNGIHTTSDQGNRDRYFSLKTDARVTTIDAHRSVLPGQWVH 60
QY 367 VAATYDGRHALYVDGTQVASSLDQSGPLNSPMAACRSLLGDSSEDCHTFRGLTL 426
DB 61 LAATYDGRMLKMLYNGAQVATSABQVGIPLQCKVLMG3--SALNHNFRGHTEHF 118
QY 427 VFWSTALPQSHFQHSQSGSGSEEAETDVLVTASPEPVNTWEPFRDEKVRLEY--LQGF 484
DB 119 SLWKVARTQREIVSDMETRGLTPRLPQLLQEWMDNVKRTWSPKQDNSQVPSNAGF 178
QY 485 EPEPEILSPQAPLPCGTVCDNVELISQYNGWPLRGEKVIROYVNICDEGLNPVSE 544
DB 179 LLD-----TNLEPPLCGTLLCDNTEVIVSYQLDSFRQPKVRRVYVNIYDHHENPVSW 234
QY 545 EQIRLQHEALNEAFSRYNISWQLSVHOVNSTLRHVVLVNCBPSKIGNDHCPECEHL 604
DB 235 QQIDFQHQQLAEAFQHNINISMELEVININSSLRHLILANCDISKIGDEKCDPECHNTL 294
QY 605 TGYDGDGCR-LQGRCVSMNRDGLCHVECNMLNDPDDGDCDQVADVKTGCFDDPS 662
DB 295 TGHGDDCRLRYPAFKKQONGVCDMDYERPNFPGGECDDPTDVTYKTCDDPSP 353

RESULT 13
ID 095L43 PRELIMINARY; PRT; 283 AA.
AC 095L43;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE Pregnancy-associated plasma protein A (Fragment).
GN PAPA-A.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA Mazerbourg S., Overgaard M.T., Oxvig C., Christiansen M.,
RA Conover C.A., Laureaud I., Vidau M., Tøsser-Klopff G., Zapf J.,

```

RA Monget P.;  
 RT "Pregnancy Associated Plasma Protein-A (PAPP-A) in ovine, bovine,  
 RT porcine and equine ovarian follicles: involvement in IGFBP-4  
 RT proteolytic degradation and mRNA expression during follicular  
 RT development";  
 RL Endocrinology 0:0-0(2001).  
 DR EMBL; AF421141; AAL16085.1; -.  
 FT NON\_TER 1  
 FT NON\_TER 283  
 SQ SEQUENCE 283 AA; 31243 MW; 866461A616F9333 CRC64;

Query Match 8.3%; Score 813.5; DB 6; Length 283;  
 Best Local Similarity 51.4%; Pred. No. 4.1e-59;  
 Matches 145; Conservative 45; Mismatches 89; Indels 3; Gaps 2;

QY 733 HEVGHVGLVHVFVGSERESCNPCKEIVSMETGULCADTATPSELCREPEPSTDT 792  
 DB 1 HEIGHSLGLHIFGIGISISQSDPCMETEPSYETGDCSTPNAPAKHFCGDPGNDT 60  
 QY 793 CGFTFPGAPFTNYSYTDNCTDNFTPNQVARMHCYLDLVYQWTSRKPPIPIPMV 852  
 DB 61 CGFTFPGAPFTNYSYTDNCTDNFTPNQVARMHCYLDLVYQWTSRKPPIPIPMV 120  
 QY 853 IGTNKSLLTHWLPPIGCVYDASGSLCGACTEDGTFROYVHTASSRRVCDSSGYWTP 912  
 DB 121 VSHTPDVMLEWEPPIIDGHFFERELGACDLCGRFLVQYAFNASSPMPCGSGHWSR 180  
 QY 913 EAVGPDPVDCPEPSLOAMSPVHLVHMNTVPCP-TEGCSLELLFQHPVQADTLTLMV 971  
 DB 181 EAGHPDVEQPCSSVATMSPNSAVNPHVTPACPEQCHLDRFRPLVPEESLTVMVT 240  
 QY 972 --SEFMESQVLFTEILLNKESVHLGPIIDTFCPIPTIKL 1011  
 DB 241 FVSTDMDSGAVNDIKLGVSGNNISLGPGVFCDVPLTIKL 282

## RESULT 14

Q95L44 PRELIMINARY; PRT; 246 AA.  
 AC 095L44;  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE Pregnancy-associated plasma protein-A (Fragment).  
 GN PAPP-A.  
 OS Ovis aries (Sheep).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Caprinae; Ovis.  
 OX NCBI\_TaxID=9940;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Mazerbourg S., Overgaard M.T., Oxvig C., Christensen M.,  
 RA Conover C.A., Laurendan I., Vidau M., Toeser-Klopp G., Zapf J.,  
 RA Monget P.;  
 RT "Pregnancy Associated Plasma Protein-A (PAPP-A) in ovine, bovine,  
 RT porcine and equine ovarian follicles: involvement in IGFBP-4  
 RT proteolytic degradation and mRNA expression during follicular  
 RT development";  
 RL Endocrinology 0:0-0(2001).  
 DR EMBL; AF421140; AAL16084.1; -.  
 FT NON\_TER 1  
 FT NON\_TER 246  
 SQ SEQUENCE 246 AA; 27275 MW; 1BFCL1D34C4E9F07 CRC64;

Query Match 7.6%; Score 752.5; DB 6; Length 246;  
 Best Local Similarity 53.3%; Pred. No. 3.9e-54;  
 Matches 128; Conservative 38; Mismatches 73; Indels 1; Gaps 1;

QY 733 HEVGHVGLVHVFVGSERESCNPCKEIVSMETGULCADTATPSELCREPEPSTDT 792  
 DB 1 HEIGHSLGLHIFGIGISISQSDPCMETEPSYETGDCSTPNAPAKHFCGDPGNDT 60

QY 793 CGFTFPGAPFTNYSYTDNCTDNFTPNQVARMHCYLDLVYQWTSRKPPIPIPMV 852  
 DB 61 CGFTFPGAPFTNYSYTDNCTDNFTPNQVARMHCYLDLVYQWTSRKPPIPIPMV 120  
 QY 853 IGTNKSLLTHWLPPIGCVYDASGSLCGACTEDGTFROYVHTASSRRVCDSSGYWTP 912  
 DB 121 VSHTPDVMLEWEPPIIDGHFFERELGACDLCGRFLVQYAFNASSPMPCGSGHWSR 180  
 QY 913 EAVGPDPVDCPEPSLOAMSPVHLVHMNTVPCP-TEGCSLELLFQHPVQADTLTLMV 971  
 DB 181 EAGHPDVEQPCSSVATMSPNSAVNPHVTPACPEQCHLDRFRPLVPEESLTVMVT 240

## RESULT 15

Q95L42 PRELIMINARY; PRT; 213 AA.  
 AC 095L42;  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE Pregnancy-associated plasma protein-A (Fragment).  
 GN PAPP-A.  
 OS Sus scrofa (Pig).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 OX NCBI\_TaxID=9823;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Mazerbourg S., Overgaard M.T., Oxvig C., Christensen M.,  
 RA Conover C.A., Laurendan I., Vidau M., Toeser-Klopp G., Zapf J.,  
 RA Monget P.;  
 RT "Pregnancy Associated Plasma Protein-A (PAPP-A) in ovine, bovine,  
 RT porcine and equine ovarian follicles: involvement in IGFBP-4  
 RT proteolytic degradation and mRNA expression during follicular  
 RT development";  
 RL Endocrinology 0:0-0(2001).  
 DR EMBL; AF421142; AAL16086.1; -.  
 FT NON\_TER 1  
 FT NON\_TER 213  
 SQ SEQUENCE 213 AA; 23684 MW; D1205E8445324A8C CRC64;

Query Match 6.4%; Score 635.5; DB 6; Length 213;  
 Best Local Similarity 51.6%; Pred. No. 1.8e-44;  
 Matches 110; Conservative 30; Mismatches 72; Indels 1; Gaps 1;

QY 758 CKETPSMETGDCIADTATPSELCREPEPSTDTGFTFPGAPFTNYSYTDNCTDN 817  
 DB 1 CKETPSMETGDCIADTATPSELCREPEPSTDTGFTFPGAPFTNYSYTDNCTDN 60  
 QY 818 FTNNOVARMHCYLDLVYQWTSRKPPIPIPMVIGQTNKSLTHWLPPIGCVYDAS 877  
 DB 61 FTNNOVARMHCYLDLVYQWTSRKPPIPIPMVIGQTNKSLTHWLPPIGCVYDAS 120  
 QY 878 GSLCGACTEDGTFROYVHTASSRRVCDSSGYWTPPEAVGPDPVDCPEPSLOAMSPVHL 937  
 DB 121 GSNACDLCEBRILIVQYAFNASSPMPCGSGHWSRPREAEGHPDVEQPCSSVATMSPNSAV 180  
 QY 938 YHNMNTVPCP-TEGCSLELLFQHPVQADTLTLM 969  
 DB 181 NPHVTPACPEQCHLDRFRPLVPEESLTVMVT 213

Search completed: January 2, 2004, 16:04:28  
 Job time : 64 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2004, 16:05:42 ; Search time 50 Seconds

(without alignments)  
7186.968 Million cell updates/sec

Title: US-09-983-025a-2

Perfect score: 9856  
Sequence: 1 MMCLIKIRISLIALAGMALC.....AADCDDECTCRPKAEENG 1791

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 733937 seqs, 200641211 residues

Total number of hits satisfying chosen parameters: 733937

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Published Applications AA:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCR\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/PCRUS\_PUBCOMB.pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
8: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*  
17: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9856	100.0	1791	11	US-09-983-025-2
2	9836	99.8	1791	10	US-09-983-025-2
3	9507	96.5	1770	10	US-09-827-998-10
4	7363	74.7	1385	10	US-09-827-998-16
5	3916.5	39.7	1627	11	US-09-983-025-25
6	3916.5	39.7	1627	12	US-10-295-027-663
7	1086	11.0	192	9	US-09-864-761-34265
8	383	3.9	70	9	US-09-864-761-34265
9	360.5	3.7	165	9	US-09-864-761-42873
10	346.5	3.5	3567	12	US-10-028-248A-47
11	336.5	3.4	3594	10	US-09-911-842-4
12	336.5	3.4	3594	10	US-10-150-821-4
13	334.5	3.4	3557	12	US-10-295-027-430
14	334.5	3.4	3557	12	US-10-295-027-1297
15	332.5	3.4	3568	12	US-10-028-248A-8

16	332.5	3.4	3570	12	US-10-028-248A-6	Sequence 6, Appl1
17	330.5	3.4	3571	10	US-09-911-842-2	Sequence 2, Appl1
18	330.5	3.4	3571	14	US-10-150-821-2	Sequence 2, Appl1
19	324	3.3	63	9	US-09-864-761-34262	Sequence 34262, A
20	292.5	3.0	2489	10	US-09-911-842-5	Sequence 5, Appl1
21	292.5	3.0	2489	10	US-10-150-821-5	Sequence 5, Appl1
22	283	2.9	1139	9	US-09-764-853-634	Sequence 634, App
23	283	2.9	1139	9	US-09-764-853-634	Sequence 204, App
24	283	2.9	1139	11	US-09-764-881-99	Sequence 99, Appl
25	283	2.9	1139	15	US-10-073-865-102	Sequence 102, App
26	278.5	2.8	2050	10	US-09-898-570-22	Sequence 22, Appl
27	278.5	2.8	2050	11	US-09-839-446-22	Sequence 22, Appl
28	271	2.7	669	12	US-10-028-248A-49	Sequence 49, Appl
29	270	2.7	601	12	US-10-028-248A-50	Sequence 50, Appl
30	249	2.5	830	15	US-10-020-141-10	Sequence 10, Appl
31	238	2.4	487	9	US-09-764-853-816	Sequence 816, App
32	238	2.4	487	9	US-09-764-898-276	Sequence 276, App
33	238	2.4	497	11	US-09-764-881-157	Sequence 157, App
34	238	2.4	497	15	US-10-073-865-125	Sequence 125, App
35	230.5	2.3	610	11	US-09-802-640-36	Sequence 36, Appl
36	230.5	2.3	610	12	US-10-021-660-122	Sequence 122, App
37	230.5	2.3	610	12	US-10-234-041-2	Sequence 2, Appl1
38	230.5	2.3	610	12	US-10-295-027-28	Sequence 28, Appl1
39	230.5	2.3	610	12	US-10-295-027-730	Sequence 730, App
40	230.5	2.3	610	12	US-10-295-027-855	Sequence 855, App
41	230.5	2.3	610	12	US-10-295-027-938	Sequence 938, App
42	230.5	2.3	610	15	US-10-205-823-357	Sequence 357, App
43	228.5	2.3	406	10	US-09-898-570-20	Sequence 20, Appl
44	228.5	2.3	406	11	US-09-839-446-20	Sequence 20, Appl
45	225	2.3	1394	12	US-10-116-275-261	Sequence 261, App

# ALIGNMENTS

RESULT 1  
US-09-983-025-2  
; Sequence 2, Application US/09983025  
; Publication No. US20030124529A1  
GENERAL INFORMATION:  
; APPLICANT: OXVIG, Claus  
; APPLICANT: OVERGARD, Michael T.  
; TITLE OF INVENTION: PRESENCE-ASSOCIATED PLASMA PROTEIN-A2 (PAP-A2)  
; FILE REFERENCE: OXVIG-1A  
; CURRENT APPLICATION NUMBER: US/09/983, 025  
; CURRENT FILING DATE: 2001-10-22  
; PRIOR APPLICATION NUMBER: US 60/241,840  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: DK PA 2000 01571  
; NUMBER OF SEQ ID NOS: 25  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2  
; LENGTH: 1791  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(66)  
; OTHER INFORMATION: prepro part of PAP-A2  
; NAME/KEY: misc\_feature  
; LOCATION: (67)..(699)  
; OTHER INFORMATION: pro part of PAP-A2  
US-09-983-025-2

Query Match 100.0%; Score 9856; DB 11; Length 1791;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1791; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MMCLIKIRISLIALAGMALCSANSEIGWTRKSLVEREHINQVILEGECRWLCAKVRPR 60  
DB 1 MMCLIKIRISLIALAGMALCSANSEIGWTRKSLVEREHINQVILEGECRWLCAKVRPR 60

61 ASPOHLLFGVYPSRAGNYLRPYVGEORIEHHTGRSKPTEGNNAVSLVPEDLTENPAGLRG 120  
61 ASPHLLFGVYPSRAGNYLRPYVGEORIEHHTGRSKPTEGNNAVSLVPEDLTENPAGLRG 120  
121 AVEBPAPWVDSPIGOSELLGDDDAYIGNORSKESLGEAGIQKGSAMAAITTTAIFTTL 180  
121 AVEBPAPWVDSPIGOSELLGDDDAYIGNORSKESLGEAGIQKGSAMAAITTTAIFTTL 180  
181 NEPEPEORRGMAKSRORROYWKRRARNDGSDGSISSHFOPWPKSLKRRKKSPPRESN 240  
181 NEPEPEORRGMAKSRORROYWKRRARNDGSDGSISSHFOPWPKSLKRRKKSPPRESN 240  
241 QNGEGSYREAEFTNSQVGLPILYFSGRERLLRPEVLAIEIPREAFTEAMVPEEGON 300  
241 QNGEGSYREAEFTNSQVGLPILYFSGRERLLRPEVLAIEIPREAFTEAMVPEEGON 300  
241 QNGEGSYREAEFTNSQVGLPILYFSGRERLLRPEVLAIEIPREAFTEAMVPEEGON 300  
301 NPATIAGVFDCSHITVSDKGMALGIRSGDKGKRDARFFSLCTDRVKKATILISHSRQ 360  
301 NPATIAGVFDCSHITVSDKGMALGIRSGDKGKRDARFFSLCTDRVKKATILISHSRQ 360  
361 PGTWTHAATYDGRHMAIYVDGTVOASSLDGSGPLNSPFMASCRSLIGDSSSDGHYFR 420  
361 PGTWTHAATYDGRHMAIYVDGTVOASSLDGSGPLNSPFMASCRSLIGDSSSDGHYFR 420  
421 GHLGTLVFWSTALPQSHFOHSSOHSSGEEATDVLTAFFEPVNTWVPEFDEKXPRLV 480  
421 GHLGTLVFWSTALPQSHFOHSSOHSSGEEATDVLTAFFEPVNTWVPEFDEKXPRLV 480  
481 LOGEPEPEELISPIOPICGQTVCDNVELISQYNGYMPLEBEKTYRYOVANI CDDEGLNP 540  
481 LOGEPEPEELISPIOPICGQTVCDNVELISQYNGYMPLEBEKTYRYOVANI CDDEGLNP 540  
481 LOGEPEPEELISPIOPICGQTVCDNVELISQYNGYMPLEBEKTYRYOVANI CDDEGLNP 540  
541 IYSEBQIRLOHEALNEAFSRYNISMOISVHOVHNSLTLEHRYVLANCCEPSKIGNHCDDEC 600  
541 IYSEBQIRLOHEALNEAFSRYNISMOISVHOVHNSLTLEHRYVLANCCEPSKIGNHCDDEC 600  
541 IYSEBQIRLOHEALNEAFSRYNISMOISVHOVHNSLTLEHRYVLANCCEPSKIGNHCDDEC 600  
601 EHPITGYDGDGCRLOGRCYSWNRBDGLCHECNMMLNFDGDCDDPOVADVRKTCFEPD 660  
601 EHPITGYDGDGCRLOGRCYSWNRBDGLCHECNMMLNFDGDCDDPOVADVRKTCFEPD 660  
601 EHPITGYDGDGCRLOGRCYSWNRBDGLCHECNMMLNFDGDCDDPOVADVRKTCFEPD 660  
661 SPKAYNSVKELKEALQUNSTHFLNTYASSVREDLAGAATWPMDKDAVTLGGIVLSPA 720  
661 SPKAYNSVKELKEALQUNSTHFLNTYASSVREDLAGAATWPMDKDAVTLGGIVLSPA 720  
661 SPKAYNSVKELKEALQUNSTHFLNTYASSVREDLAGAATWPMDKDAVTLGGIVLSPA 720  
721 YXGPHGHTDMMIHEVHGLGLYHVEKGVSEBSCNDPCKETVPSEMETDLCADTAPEPKS 780  
721 YXGPHGHTDMMIHEVHGLGLYHVEKGVSEBSCNDPCKETVPSEMETDLCADTAPEPKS 780  
721 YXGPHGHTDMMIHEVHGLGLYHVEKGVSEBSCNDPCKETVPSEMETDLCADTAPEPKS 780  
781 ELCEPEPEPTSDTCGFTFRPGAPFTNYSYTDNCTDNFTPNQVARMHCYLDLVYQOMTES 840  
781 ELCEPEPEPTSDTCGFTFRPGAPFTNYSYTDNCTDNFTPNQVARMHCYLDLVYQOMTES 840  
781 ELCEPEPEPTSDTCGFTFRPGAPFTNYSYTDNCTDNFTPNQVARMHCYLDLVYQOMTES 840  
841 RKPTPIPIPPWVIGQTNKSLTIHMLPISGVVYDRASSGLCGACTEDGTFRQYVHTASSR 900  
841 RKPTPIPIPPWVIGQTNKSLTIHMLPISGVVYDRASSGLCGACTEDGTFRQYVHTASSR 900  
841 RKPTPIPIPPWVIGQTNKSLTIHMLPISGVVYDRASSGLCGACTEDGTFRQYVHTASSR 900  
901 RVCSGSGWPEBEAVGPDPVOPCEPSLOAMSPEVHLVHMNTVPCPEGSGLELTFPH 960  
901 RVCSGSGWPEBEAVGPDPVOPCEPSLOAMSPEVHLVHMNTVPCPEGSGLELTFPH 960  
901 RVCSGSGWPEBEAVGPDPVOPCEPSLOAMSPEVHLVHMNTVPCPEGSGLELTFPH 960  
961 VQADTLTLMVTSFFNESSQVLFDTIELLENKESVHGLDFTFCOILPLTIKLVHVGKSVGV 1020  
961 VQADTLTLMVTSFFNESSQVLFDTIELLENKESVHGLDFTFCOILPLTIKLVHVGKSVGV 1020  
961 VQADTLTLMVTSFFNESSQVLFDTIELLENKESVHGLDFTFCOILPLTIKLVHVGKSVGV 1020  
1021 KYVTFDEIEITDALLISQPHSPLCSGCRPVRYOVLNDPPRASGLPVVVTSHRKTDFVE 1080  
1021 KYVTFDEIEITDALLISQPHSPLCSGCRPVRYOVLNDPPRASGLPVVVTSHRKTDFVE 1080  
1021 KYVTFDEIEITDALLISQPHSPLCSGCRPVRYOVLNDPPRASGLPVVVTSHRKTDFVE 1080  
1081 VTPGMYOYVLAABGELGEASPLNIIHGAAPYCGDDKVSERLCEEDDDDLVSGDCS 1140  
1081 VTPGMYOYVLAABGELGEASPLNIIHGAAPYCGDDKVSERLCEEDDDDLVSGDCS 1140  
1081 VTPGMYOYVLAABGELGEASPLNIIHGAAPYCGDDKVSERLCEEDDDDLVSGDCS 1140  
1141 KYCELEBGFNCVGEPSLICYMEGDGICPPEFKTISVDCGIYTPKGYLDQWATRAYSSHE 1200

1141 KYCELEBGFNCVGEPSLICYMEGDGICPPEFKTISVDCGIYTPKGYLDQWATRAYSSHE 1200  
1201 DKKCCPVSLVTGEPHSLICTSYHBDLPNNRPLTGMFPFVANSNEMTODDRSEOGESLKE 1260  
1201 DKKCCPVSLVTGEPHSLICTSYHBDLPNNRPLTGMFPFVANSNEMTODDRSEOGESLKE 1260  
1261 DEWMLKVCNRPGBARAIPIFLTTDGLVPEHOOPTVTLVLTVRGNSHSLGTYSQCH 1320  
1261 DEWMLKVCNRPGBARAIPIFLTTDGLVPEHOOPTVTLVLTVRGNSHSLGTYSQCH 1320  
1261 DEWMLKVCNRPGBARAIPIFLTTDGLVPEHOOPTVTLVLTVRGNSHSLGTYSQCH 1320  
1321 NPILINVTTHQNVLFHHTTSVLNFSFPRVIGSAVALRTSSRIGLASPNCSISEDEQNH 1380  
1321 NPILINVTTHQNVLFHHTTSVLNFSFPRVIGSAVALRTSSRIGLASPNCSISEDEQNH 1380  
1321 NPILINVTTHQNVLFHHTTSVLNFSFPRVIGSAVALRTSSRIGLASPNCSISEDEQNH 1380  
1381 QGOSCHRPCKGKDCSCLLDHADVNCISIGPLMKCAITCORGFALQASSGQYIRPM 1440  
1381 QGOSCHRPCKGKDCSCLLDHADVNCISIGPLMKCAITCORGFALQASSGQYIRPM 1440  
1381 QGOSCHRPCKGKDCSCLLDHADVNCISIGPLMKCAITCORGFALQASSGQYIRPM 1440  
1441 QKEILLTCSGSHMDONVSCLPVDCGVPDPSLVNANFSCSEGTKFLKRCGISCVPPAKQ 1500  
1441 QKEILLTCSGSHMDONVSCLPVDCGVPDPSLVNANFSCSEGTKFLKRCGISCVPPAKQ 1500  
1441 QKEILLTCSGSHMDONVSCLPVDCGVPDPSLVNANFSCSEGTKFLKRCGISCVPPAKQ 1500  
1501 GLSPMLTCLBDGLMSLPEVYCKLECDAPPIILNANILLPHCLDNDHVGITICKYCKPXY 1560  
1501 GLSPMLTCLBDGLMSLPEVYCKLECDAPPIILNANILLPHCLDNDHVGITICKYCKPXY 1560  
1501 GLSPMLTCLBDGLMSLPEVYCKLECDAPPIILNANILLPHCLDNDHVGITICKYCKPXY 1560  
1561 YVNASAGKRNKLLKIQCLEGGIWEBSGCIPIVCEPPEPVPFEGMVECTMGFSLDSQCVL 1620  
1561 YVNASAGKRNKLLKIQCLEGGIWEBSGCIPIVCEPPEPVPFEGMVECTMGFSLDSQCVL 1620  
1561 YVNASAGKRNKLLKIQCLEGGIWEBSGCIPIVCEPPEPVPFEGMVECTMGFSLDSQCVL 1620  
1621 NCNORERKLPILCTKEGLMTQEFKLCENTLOGECPPEPSSEANSYKCEOGYIGAVCSPL 1680  
1621 NCNORERKLPILCTKEGLMTQEFKLCENTLOGECPPEPSSEANSYKCEOGYIGAVCSPL 1680  
1621 NCNORERKLPILCTKEGLMTQEFKLCENTLOGECPPEPSSEANSYKCEOGYIGAVCSPL 1680  
1681 CVIPSPDPMVLPENITADTLLEHMMBPKVQSIYCTGRQWHPDVLVHCIOQCEPFGADG 1740  
1681 CVIPSPDPMVLPENITADTLLEHMMBPKVQSIYCTGRQWHPDVLVHCIOQCEPFGADG 1740  
1681 CVIPSPDPMVLPENITADTLLEHMMBPKVQSIYCTGRQWHPDVLVHCIOQCEPFGADG 1740  
1741 WCDTINNRAICHYDGGCCSSTLSSKVIIPFADCDLDECTCRDPKAEHQ 1791  
1741 WCDTINNRAICHYDGGCCSSTLSSKVIIPFADCDLDECTCRDPKAEHQ 1791  
1741 WCDTINNRAICHYDGGCCSSTLSSKVIIPFADCDLDECTCRDPKAEHQ 1791

RESULT 2  
US-09-983-98-3  
Sequence 3, Application US/09827998  
Patent No. US20020102252A1  
GENERAL INFORMATION:  
APPLICANT: Gu, Yizhong  
TITLE OF INVENTION: NOVEL ISOFORMS OF HUMAN PREGNANCY-ASSOCIATED PROTEIN E  
FILE REFERENCE: NM0087-8  
CURRENT APPLICATION NUMBER: US/09/827,998  
CURRENT FILING DATE: 2001-04-06  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
NUMBER OF SEQ ID NOS: 1881  
SOFTWARE: Neomica Sequence Listing Engine  
SEQ ID NO 3  
LENGTH: 1791  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-983-98-3

Query Match 99.8%; Score 9836; DB 10; Length 1791;  
Best Local Similarity 99.8%; Pred. No. 0; Indels 0; Gaps 0;  
Matches 1788; Conservative 1; Mismatches 2;  
1 MMCLTIIRISLAIAGALCSANSELGWTKKSLVEREHLNOVLLBGERCWLGAQVRRPR 60

```

Db      1 MMCKLILSLAILAGLALCSANSELGWTNRKSLVEREHLNOVLLEGERCMLGAKVRRPR 60
Qy      61 ASFOHHLFGVYPSRAGNYLRPYVGEBOIHHTGRSKDETEGNAVSLVPPDLTENPAGLRG 120
Db      61 ASFOHHLFGVYPSRAGNYLRPYVGEBOIHHTGRSKDETEGNAVSLVPPDLTENPAGLRG 120
Qy      121 AVEBPAPWVGDSPIQSGSELLGDDDAYLGNQSRKESLGEAGIOKGSAMAAITTTAIFPTTL 180
Db      121 AVEBPAPWVGDSPIQSGSELLGDDDAYLGNQSRKESLGEAGIOKGSAMAAITTTAIFPTTL 180
Qy      181 NEPEPELQRBGMASRRQRYWKRRADGGDSGISSHFQPMFNGSLKHKVKSPPRESN 240
Db      181 NEPEPELQRBGMASRRQRYWKRRADGGDSGISSHFQPMFNGSLKHKVKSPPRESN 240
Qy      241 ONGEGSYREAFETNSOVGLPILYFSGRRELLRLRPVLAIEIPEAFETVEAMVPEEGON 300
Db      241 ONGEGSYREAFETNSOVGLPILYFSGRRELLRLRPVLAIEIPEAFETVEAMVPEEGON 300
Qy      301 NPALIAVFNDCSHTVSDKGMALGIRSGDKGKRDAFFPSLCTDRYKATITLISHSRQ 360
Db      301 NPALIAVFNDCSHTVSDKGMALGIRSGDKGKRDAFFPSLCTDRYKATITLISHSRQ 360
Qy      361 PGTWTHAATYDGRHMAIYDGTQVASSLDQSGPLNSPFMASSCSLLIGDSSSEGHYFR 420
Db      361 PGTWTHAATYDGRHMAIYDGTQVASSLDQSGPLNSPFMASSCSLLIGDSSSEGHYFR 420
Qy      421 GHGLTVLFWGSTALPQSHFQSHSSHSGSEBEATDVLTAPEFVTEWVPFDEKXPYLEV 480
Db      421 GHGLTVLFWGSTALPQSHFQSHSSHSGSEBEATDVLTAPEFVTEWVPFDEKXPYLEV 480
Qy      481 LOGEPEPELILSPLOPLCQGTVCNDELISQYNGYPLRGEKVIYQVANI CDDEGLNP 540
Db      481 LOGEPEPELILSPLOPLCQGTVCNDELISQYNGYPLRGEKVIYQVANI CDDEGLNP 540
Qy      541 IVSEBOQLRLOHEALNEAFSRYNISWQISVHOVNSTRLHRVYLVNCEPSKIGNHCPDEC 600
Db      541 IVSEBOQLRLOHEALNEAFSRYNISWQISVHOVNSTRLHRVYLVNCEPSKIGNHCPDEC 600
Qy      601 EHPITGVDGDCRLQGRCSYNNRRDGLCHVECNMMLDPDGDCCDQVAVRKTCTCPD 660
Db      601 EHPITGVDGDCRLQGRCSYNNRRDGLCHVECNMMLDPDGDCCDQVAVRKTCTCPD 660
Qy      661 SPKRAVMSVKEKALQJNSTHFLNIYFASVREDLAGAATPMDKXAVTHLGGIIVLSPA 720
Db      661 SPKRAVMSVKEKALQJNSTHFLNIYFASVREDLAGAATPMDKXAVTHLGGIIVLSPA 720
Qy      721 YGMPGHTDTMIEHGVGLIYHVFKGVSERESCNDCCKETVPSMETGDLCAOTAPTPKS 780
Db      721 YGMPGHTDTMIEHGVGLIYHVFKGVSERESCNDCCKETVPSMETGDLCAOTAPTPKS 780
Qy      781 ELCEPEPEPTDTCGTFRPGAPPTNNWMSYTDNCTDNFTPOVARMCIYDLVYQOMTES 840
Db      781 ELCEPEPEPTDTCGTFRPGAPPTNNWMSYTDNCTDNFTPOVARMCIYDLVYQOMTES 840
Qy      841 RKPTPIPIPPVIGQTKSLITIMLPRISGVYDRASSGLGACTEBGTFRQYVHTASSR 900
Db      841 RKPTPIPIPPVIGQTKSLITIMLPRISGVYDRASSGLGACTEBGTFRQYVHTASSR 900
Qy      901 RVCDSGMYTPEEAVPDVQDCEPSLOAMSPEVLYHNMNTVPCPTBGCSLELLQHP 960
Db      901 RVCDSGMYTPEEAVPDVQDCEPSLOAMSPEVLYHNMNTVPCPTBGCSLELLQHP 960
Qy      961 VQADTLTLMWTSFMESQVLFTEILLNKESVHLGPDLCOIPLTIKHLVGVSGV 1020
Db      961 VQADTLTLMWTSFMESQVLFTEILLNKESVHLGPDLCOIPLTIKHLVGVSGV 1020
Qy      1021 KVTYFDERIEIDALLTSQPHSPICSGCRPVRYOVLADPPASGLPVVTHSHKFTDVE 1080
Db      1021 KVTYFDERIEIDALLTSQPHSPICSGCRPVRYOVLADPPASGLPVVTHSHKFTDVE 1080
Qy      1081 VTPGQMTQYOVYLABGELGASAPPLNHIHGAFCGDKVSERLGEBCDDGLVSGDGS 1140
Db      1081 VTPGQMTQYOVYLABGELGASAPPLNHIHGAFCGDKVSERLGEBCDDGLVSGDGS 1140

```

```

Qy      1141 KVELEBEGFNCVGEPSLCTYMEGDGICEPFRKTSIVDCGIYTPKGYLDQMATRAYSHE 1200
Db      1141 KVELEBEGFNCVGEPSLCTYMEGDGICEPFRKTSIVDCGIYTPKGYLDQMATRAYSHE 1200
Qy      1201 DKKCPVSLVTGEPHSLICTSYHDPJPNRPLTGMFPVCASENETODRSEQPEGLKKE 1260
Db      1201 DKKCPVSLVTGEPHSLICTSYHDPJPNRPLTGMFPVCASENETODRSEQPEGLKKE 1260
Qy      1261 DEWMLKCFNRPEBARAIFLITTDGLVPEHQPTVTLVLTVRGNSHSLGTYGSLCOH 1320
Db      1261 DEWMLKCFNRPEBARAIFLITTDGLVPEHQPTVTLVLTVRGNSHSLGTYGSLCOH 1320
Qy      1321 NPLIINTVHTQNVLFHTTTSVLNFSAPRGISAVALTSSRIGLSAPNSCISEDEGONH 1380
Db      1321 NPLIINTVHTQNVLFHTTTSVLNFSAPRGISAVALTSSRIGLSAPNSCISEDEGONH 1380
Qy      1381 QGOSCIHRPCGKQDCPSLLDHDADVNCISIGPLMKCAITTCORFALQASSGOYIRPM 1440
Db      1381 QGOSCIHRPCGKQDCPSLLDHDADVNCISIGPLMKCAITTCORFALQASSGOYIRPM 1440
Qy      1441 QKSIILLTSSGMDQNVSCLPVDCGVPDPSLVNYANFSCSEGTIKLRCSISCVPAKLQ 1500
Db      1441 QKSIILLTSSGMDQNVSCLPVDCGVPDPSLVNYANFSCSEGTIKLRCSISCVPAKLQ 1500
Qy      1501 GLSPMLTCLBEDGLMSLPEVYCKLECDAPITILNANLLPCLDNDHVGITCKYECKRGY 1560
Db      1501 GLSPMLTCLBEDGLMSLPEVYCKLECDAPITILNANLLPCLDNDHVGITCKYECKRGY 1560
Qy      1561 YVASEAGKVRNKLKLIQCLEGGIWEQSGCIPVCEPBPVPEFGMYECTNGFSLDSQCVL 1620
Db      1561 YVASEAGKVRNKLKLIQCLEGGIWEQSGCIPVCEPBPVPEFGMYECTNGFSLDSQCVL 1620
Qy      1621 NCMQERKLPILCTKESLWQERKLCENLOGECPPEPSELSNVEYKKEOGYIGAVCSPL 1680
Db      1621 NCMQERKLPILCTKESLWQERKLCENLOGECPPEPSELSNVEYKKEOGYIGAVCSPL 1680
Qy      1681 CVIPSPDPVMLPENITADTLEHMMEPYKQSIYCTGRQWHPDVLVHICIQSCFPQADG 1740
Db      1681 CVIPSPDPVMLPENITADTLEHMMEPYKQSIYCTGRQWHPDVLVHICIQSCFPQADG 1740
Qy      1741 MCDTINNRAACHYDGGDCSSSTLSSKKVIPAADCDLDECTCRDPAKENQ 1791
Db      1741 MCDTINNRAACHYDGGDCSSSTLSSKKVIPAADCDLDECTCRDPAKENQ 1791

```

```

RESULT 3
US-09-827-998-10
: Sequence 10, Application US/09827998
: Patent No. US20020102252A1
: GENERAL INFORMATION:
: APPLICANT: Gu, Yizhong
: APPLICANT: Shannon, Mark
: TITLE OF INVENTION: NOVEL ISOFORMS OF HUMAN PREGNANCY-ASSOCIATED PROTEIN E
: FILE REFERENCE: MDHMR-8
: CURRENT APPLICATION NUMBER: US/09/827, 998
: CURRENT FILING DATE: 2001-04-06
: PRIOR APPLICATION NUMBER: US 60/207, 456
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: US 60/236, 359
: PRIOR FILING DATE: 2000-09-27
: NUMBER OF SEQ ID NOS: 1881
: SOFTWARE: Aeonica Sequence Listing Engine
: SEQ ID NO 10
: LENGTH: 1770
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-827-998-10

```

```

Query Match      96.5%; Score 9507; DB 10; Length 1770;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1732; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

Qy 1 MMCKLIRISLAIAGNALCSANSELGWTTRKSLVEREHLNOVLLEGERCWLGAQVRR 60
Db 1 MMCKLIRISLAIAGNALCSANSELGWTTRKSLVEREHLNOVLLEGERCWLGAQVRR 60
Qy 61 ASPQHLFGVYPSAAGNYLRPYVGEQETIHTGSKPTBGNANVSLVPPDLTENPAGIRG 120
Db 61 ASPQHLFGVYPSAAGNYLRPYVGEQETIHTGSKPTBGNANVSLVPPDLTENPAGIRG 120
Qy 121 AVEEPAAPWVDSPIGSELLGDDDAVLGNQSKESLGEAGIQGSAAMAATTITAIPTTL 180
Db 121 AVEEPAAPWVDSPIGSELLGDDDAVLGNQSKESLGEAGIQGSAAMAATTITAIPTTL 180
Qy 181 NEPEPEORRGMAKSRORQYWKRAEDQGDGSISSHFOPMKSLKGRVKSPREESN 240
Db 181 NEPEPEORRGMAKSRORQYWKRAEDQGDGSISSHFOPMKSLKGRVKSPREESN 240
Qy 241 QNGEGSREARETNSQVGLPILYFSGRERLLRPEVLAIIPREAFVEMVKEGQN 300
Db 241 QNGEGSREARETNSQVGLPILYFSGRERLLRPEVLAIIPREAFVEMVKEGQN 300
Qy 301 NPATIAGVFDCSHTVSDKGMALGIRSGDKGKDPARFPSLCTDRVKKATITLISHRYQ 360
Db 301 NPATIAGVFDCSHTVSDKGMALGIRSGDKGKDPARFPSLCTDRVKKATITLISHRYQ 360
Qy 361 PGITWVAATYDGRHMLYVDGTQVASSLDQSGPLNSPFMASCSRLILGDSSEDEGHYR 420
Db 361 PGITWVAATYDGRHMLYVDGTQVASSLDQSGPLNSPFMASCSRLILGDSSEDEGHYR 420
Qy 421 GHLGLTYFWSTALQSHQSHSSGSEBEATDVLTAISFEPVNTENVPPFDEKXPRLV 480
Db 421 GHLGLTYFWSTALQSHQSHSSGSEBEATDVLTAISFEPVNTENVPPFDEKXPRLV 480
Qy 481 LQGEPEPEELSPLOPLCGQTVCDNVELISQYNGWYPLRGEKYIRYOVNICDEGLNP 540
Db 481 LQGEPEPEELSPLOPLCGQTVCDNVELISQYNGWYPLRGEKYIRYOVNICDEGLNP 540
Qy 541 IVSEBQIRLQHEALNEAFSRYNISWQSVHQVHNSLTHRRVVLVNCBPSKIGNDHCEPC 600
Db 541 IVSEBQIRLQHEALNEAFSRYNISWQSVHQVHNSLTHRRVVLVNCBPSKIGNDHCEPC 600
Qy 601 EHPITGYDGDRCRLOGRCYSMNRDGLCHVECNMMLNFDGDCDDCPVAVNRKTCFPPD 660
Db 601 EHPITGYDGDRCRLOGRCYSMNRDGLCHVECNMMLNFDGDCDDCPVAVNRKTCFPPD 660
Qy 661 SPKXAVSVKELKEALQNSTHFLNIYFASSVREDLAGAATWPMKDXVTHLGGIVLSPA 720
Db 661 SPKXAVSVKELKEALQNSTHFLNIYFASSVREDLAGAATWPMKDXVTHLGGIVLSPA 720
Qy 721 YYGMPGHTDTHIEVGHVLAGLVHVFQVSVRESQNDPKETVPSMETGDLCAATAPTPKS 780
Db 721 YYGMPGHTDTHIEVGHVLAGLVHVFQVSVRESQNDPKETVPSMETGDLCAATAPTPKS 780
Qy 781 ELCEPEPTSDTCGTRFPGARPTNYMSYTDNCTDNFTPNQVAMHCYLDLVYQOMTES 840
Db 781 ELCEPEPTSDTCGTRFPGARPTNYMSYTDNCTDNFTPNQVAMHCYLDLVYQOMTES 840
Qy 841 RKPIPIPIPVWIGQTKSLTIHMLPRISGVVYBRASGLGACEDTCTFROYVHTASSR 900
Db 841 RKPIPIPIPVWIGQTKSLTIHMLPRISGVVYBRASGLGACEDTCTFROYVHTASSR 900
Qy 901 RVCSGSGWTPBEBAVGPDPVQPCPSLQAMSPEVHLTHMMNTVPCPTGEGSLLELFOHP 960
Db 901 RVCSGSGWTPBEBAVGPDPVQPCPSLQAMSPEVHLTHMMNTVPCPTGEGSLLELFOHP 960
Qy 961 VQADTLITMTVSFMESQVLPDTEILLENKESVHLGLDTPCDIPLTIKHLVHGKVS 1020
Db 961 VQADTLITMTVSFMESQVLPDTEILLENKESVHLGLDTPCDIPLTIKHLVHGKVS 1020
Qy 1021 KVTYFDEEIEIDALLTSQHSPLCSGCRPVRYOVLBPAPASGIPVYVTHSHRKTIVE 1080
Db 1021 KVTYFDEEIEIDALLTSQHSPLCSGCRPVRYOVLBPAPASGIPVYVTHSHRKTIVE 1080
Qy 1081 VTPGOMYOVLAEAGELGEASPLNHIHGAFCYCGDKVSERLGECDSDILVSGDCS 1140

```

```

Db 1081 VTPGOMYOVLAEAGELGEASPLNHIHGAFCYCGDKVSERLGECDSDILVSGDCS 1140
Qy 1141 KVCELEGFNCVCEPSLCVNYEGDTCPEPERKTSYVDCGITYPKGLDMQATRAYSHE 1200
Db 1141 KVCELEGFNCVCEPSLCVNYEGDTCPEPERKTSYVDCGITYPKGLDMQATRAYSHE 1200
Qy 1201 DKKCPVSLVTGEBHSILCTSYNPDLPNHRPLTGMFPVASENETODDRSEOGESLJKE 1260
Db 1201 DKKCPVSLVTGEBHSILCTSYNPDLPNHRPLTGMFPVASENETODDRSEOGESLJKE 1260
Qy 1261 DEWVLKVCENRPGBARAIFLITTDGLVGEHQOPTVTLVLTVDYRGSNLSLGTGLSCQH 1320
Db 1261 DEWVLKVCENRPGBARAIFLITTDGLVGEHQOPTVTLVLTVDYRGSNLSLGTGLSCQH 1320
Qy 1321 NPILINTHONVLFHHTTVLNFSSPRVIGASVARTSSRIGLSAPSNCISBDEQNH 1380
Db 1321 NPILINTHONVLFHHTTVLNFSSPRVIGASVARTSSRIGLSAPSNCISBDEQNH 1380
Qy 1381 QGSCIRPCGKODSCPSLLDHDVYNCTSIGGLMKCAITCORGFALQASSGQYIRPM 1440
Db 1381 QGSCIRPCGKODSCPSLLDHDVYNCTSIGGLMKCAITCORGFALQASSGQYIRPM 1440
Qy 1441 QKEILTCSSGHWQNVSCLPVDCGVPDPSLVNYANFSCSEGTFLKRCISICVPAKIQ 1500
Db 1441 QKEILTCSSGHWQNVSCLPVDCGVPDPSLVNYANFSCSEGTFLKRCISICVPAKIQ 1500
Qy 1501 GLSPWLTCLEBDGLMSLEBYVCKLECDAPITILNANLLPHCLDNHNVGTICRYECKRGY 1560
Db 1501 GLSPWLTCLEBDGLMSLEBYVCKLECDAPITILNANLLPHCLDNHNVGTICRYECKRGY 1560
Qy 1561 YVASEAGKRYNKLKIQCELGIGWEGSCIPVYCEPPPVFEGMYECTGFSLDSQCVL 1620
Db 1561 YVASEAGKRYNKLKIQCELGIGWEGSCIPVYCEPPPVFEGMYECTGFSLDSQCVL 1620
Qy 1621 NCNGEREKLPILCTKEGLMTQEFKLCENLQGECPRPSELSNVYKCEGIGIGAVCSPL 1680
Db 1621 NCNGEREKLPILCTKEGLMTQEFKLCENLQGECPRPSELSNVYKCEGIGIGAVCSPL 1680
Qy 1681 CVTPSPDPVMLPENITADTLEHMMEPYKQSYVCTGSRQHPVPLVHCIOSE 1734
Db 1681 CVTPSPDPVMLPENITADTLEHMMEPYKQSYVCTGSRQHPVPLVHCIOSE 1734

```

## RESULT 4

```

US-09-982-998-16
; Sequence 16, Application US/09827998
; Patent No. US20020102252A1
; GENERAL INFORMATION:
; APPLICANT: Shannon, Mark
; TITLE OF INVENTION: NOVEL ISOFORMS OF HUMAN PREGNANCY-ASSOCIATED PROTEIN E
; FILE REFERENCE: MDHMOF-8
; CURRENT APPLICATION NUMBER: US/09/827,998
; PRIOR APPLICATION NUMBER: 2001-04-06
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 1881
; SOFTWARE: Aeomica Sequence Listing Engine
; SEQ ID NO 16
; LENGTH: 1385
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-827-998-16

```

```

Query Match 74.7%; Score 7363; DB 10; Length 1385;
Best Local Similarity 77.0%; Pred. No. 0;
Matches 1379; Conservative 0; Mismatches 6; Indels 406; Gaps 1;
Qy 1 MMCKLIRISLAIAGNALCSANSELGWTTRKSLVEREHLNOVLLEGERCWLGAQVRR 60

```

Db 1 MMCLKILRISLAILAGMALCSANSELGWTNRKSLVEREHLNQLLEGERCWLGAIVRRPR 60  
 Qy 61 ASQOHHLFGYVPRRANVYLRPVYVGEHOLHTHTGRSKPTTEGNVSLVPPDLTENPAGLRG 120  
 Db 61 ASQOHHLFGYVPRRANVYLRPVYVGEHOLHTHTGRSKPTTEGNVSLVPPDLTENPAGLRG 120  
 Qy 121 AVEBPAPWVGDSPICQSELLGDDAYLGNQSKESLGEAGIQKSGAMATTTTALFTTL 180  
 Db 121 AVEBPAPWVGDSPICQSELLGDDAYLGNQSKESLGEAGIQKSGAMATTTTALFTTL 180  
 Qy 181 NEKPETORRGMAKSRORQVWKRABDQGDGSGISHFQPMKSHLKKRVKSPREBN 240  
 Db 181 NEKPETORRGMAKSRORQVWKRABDQGDGSGISHFQPMKSHLKKRVKSPREBN 240  
 Qy 241 ONGEGSYREAFETNSQVGLPIYFSGRERILLRPEVLAIPREAFTEAVAWKPEGGON 300  
 Db 241 ONGEGSYREAFETNSQVGLPIYFSGRERILLRPEVLAIPREAFTEAVAWKPEGGON 300  
 Qy 301 NPALIAVFNDCSHTVSDKGMALGIRGKOKKRDANFPFSLCTDRVKKATILISHRYQ 360  
 Db 301 NPALIAVFNDCSHTVSDKGMALGIRGKOKKRDANFPFSLCTDRVKKATILISHRYQ 360  
 Qy 361 PGTHTAAATYDGRHMLLYVDGTQVASSLDQSGPLNSPFMASCSLLLGSDSSEGHYFR 420  
 Db 361 PGTHTAAATYDGRHMLLYVDGTQVASSLDQSGPLNSPFMASCSLLLGSDSSEGHYFR 420  
 Qy 421 GHLGLVFWSTALPQSHFQHSSOHSGHEEATDLVLTASFEPVNTFVWPFREDEKYPRLV 480  
 Db 421 GHLGLVFWSTALPQSHFQHSSOHSGHEEATDLVLTASFEPVNTFVWPFREDEKYPRLV 480  
 Qy 307 307 306  
 Db 307 307 306  
 Qy 481 LOGEPPEPILSPLOPLCGQVCDNVELISQYNGYPLRGEKVIYQVANNICDEGLNP 540  
 Db 481 LOGEPPEPILSPLOPLCGQVCDNVELISQYNGYPLRGEKVIYQVANNICDEGLNP 540  
 Qy 307 307 306  
 Db 307 307 306  
 Qy 541 IYSEQIRLQHEALNEAFGRYNI SWQLSVHQNSTLRHYVYLNCESKIGNDHCEC 600  
 Db 541 IYSEQIRLQHEALNEAFGRYNI SWQLSVHQNSTLRHYVYLNCESKIGNDHCEC 600  
 Qy 307 307 306  
 Db 307 307 306  
 Qy 601 EHPRLTGDGDCRLOGRCYSNNRDGLCHVECNMNLNPDGDCDDPOVADVRKTCFPPD 660  
 Db 601 EHPRLTGDGDCRLOGRCYSNNRDGLCHVECNMNLNPDGDCDDPOVADVRKTCFPPD 660  
 Qy 307 307 306  
 Db 307 307 306  
 Qy 661 SPKRAVMSKELKEALQNSTHFLNITYFASVREDLAGAATWPKDAVTHLGIVLSPA 720  
 Db 661 SPKRAVMSKELKEALQNSTHFLNITYFASVREDLAGAATWPKDAVTHLGIVLSPA 720  
 Qy 307 307 306  
 Db 307 307 306  
 Qy 721 YVGMPTHTDTHIEVHGLVLYHFKVGSERESNDPCKETVPSMETGDLCAVTPTRKS 780  
 Db 721 YVGMPTHTDTHIEVHGLVLYHFKVGSERESNDPCKETVPSMETGDLCAVTPTRKS 780  
 Qy 315 YVGMPTHTDTHIEVHGLVLYHFKVGSERESNDPCKETVPSMETGDLCAVTPTRKS 374  
 Db 315 YVGMPTHTDTHIEVHGLVLYHFKVGSERESNDPCKETVPSMETGDLCAVTPTRKS 374  
 Qy 781 ELCREBPPTSDTGCFTRFPAPFTNTNMYTDNCTNFTPNQVABWHCYLDLVYQOWTES 840  
 Db 781 ELCREBPPTSDTGCFTRFPAPFTNTNMYTDNCTNFTPNQVABWHCYLDLVYQOWTES 840  
 Qy 375 ELCREBPPTSDTGCFTRFPAPFTNTNMYTDNCTNFTPNQVABWHCYLDLVYQOWTES 434  
 Db 375 ELCREBPPTSDTGCFTRFPAPFTNTNMYTDNCTNFTPNQVABWHCYLDLVYQOWTES 434  
 Qy 841 RKPTPIIPMWIGQTKSLTIHMLPISGVVYDRAAGSLCGACTEDGTFROYVHTASSR 900  
 Db 841 RKPTPIIPMWIGQTKSLTIHMLPISGVVYDRAAGSLCGACTEDGTFROYVHTASSR 900  
 Qy 435 RKPTPIIPMWIGQTKSLTIHMLPISGVVYDRAAGSLCGACTEDGTFROYVHTASSR 494  
 Db 435 RKPTPIIPMWIGQTKSLTIHMLPISGVVYDRAAGSLCGACTEDGTFROYVHTASSR 494  
 Qy 901 RVCDSSGYWTPPEAVGPPVDQCEPSLOAMSPRYHLYHNMVVPCTECCSELLEFOP 960  
 Db 901 RVCDSSGYWTPPEAVGPPVDQCEPSLOAMSPRYHLYHNMVVPCTECCSELLEFOP 960  
 Qy 495 RVCDSSGYWTPPEAVGPPVDQCEPSLOAMSPRYHLYHNMVVPCTECCSELLEFOP 554  
 Db 495 RVCDSSGYWTPPEAVGPPVDQCEPSLOAMSPRYHLYHNMVVPCTECCSELLEFOP 554  
 Qy 961 VQADTLTLWTSFPMSSQVLFDEILLENKESVHLGAPLDTFCDIPLTIKLHVDGKVSQV 1020  
 Db 961 VQADTLTLWTSFPMSSQVLFDEILLENKESVHLGAPLDTFCDIPLTIKLHVDGKVSQV 1020  
 Qy 555 VQADTLTLWTSFPMSSQVLFDEILLENKESVHLGAPLDTFCDIPLTIKLHVDGKVSQV 614  
 Db 555 VQADTLTLWTSFPMSSQVLFDEILLENKESVHLGAPLDTFCDIPLTIKLHVDGKVSQV 614  
 Qy 1021 KVVTFDERIEIDALLTSQPHSPLCSGCRPVRYVLRDPPASGLPVVYTHSRKFTDVE 1080  
 Db 1021 KVVTFDERIEIDALLTSQPHSPLCSGCRPVRYVLRDPPASGLPVVYTHSRKFTDVE 1080  
 Qy 615 KVVTFDERIEIDALLTSQPHSPLCSGCRPVRYVLRDPPASGLPVVYTHSRKFTDVE 674  
 Db 615 KVVTFDERIEIDALLTSQPHSPLCSGCRPVRYVLRDPPASGLPVVYTHSRKFTDVE 674  
 Qy 1081 VTPGOMYQVTLAAGGELGASPLNTHIGAPYCGDGKVERLGECCDDGLVSDGCS 1140  
 Db 1081 VTPGOMYQVTLAAGGELGASPLNTHIGAPYCGDGKVERLGECCDDGLVSDGCS 1140  
 Qy 675 VTPGOMYQVTLAAGGELGASPLNTHIGAPYCGDGKVERLGECCDDGLVSDGCS 734  
 Db 675 VTPGOMYQVTLAAGGELGASPLNTHIGAPYCGDGKVERLGECCDDGLVSDGCS 734

Qy 1141 KYCEBERGFCVGPSPSLCYMYEGDICEPFRKTSIVDCGYTPKGYLDQMATRAYSSE 1200  
 Db 735 KYCEBERGFCVGPSPSLCYMYEGDICEPFRKTSIVDCGYTPKGYLDQMATRAYSSE 794  
 Qy 1201 DKKCPVSLVYGBEBSHLSICTSYHDLNHRPLTGWFPCVASNEOTODRSEOGESLJKE 1260  
 Db 795 DKKCPVSLVYGBEBSHLSICTSYHDLNHRPLTGWFPCVASNEOTODRSEOGESLJKE 854  
 Qy 1261 DEWMLKVCFNRRPGARAFIFLITDGLVPGHOOPVTLVLTVDKSNHSLGTYGSLSCQ 1320  
 Db 855 DEWMLKVCFNRRPGARAFIFLITDGLVPGHOOPVTLVLTVDKSNHSLGTYGSLSCQ 914  
 Qy 1321 NPLIINTVHONVLFHTTSVTLNFSRVSIGSALARTSSRIGLSAPNSCISDEBQNH 1380  
 Db 915 NPLIINTVHONVLFHTTSVTLNFSRVSIGSALARTSSRIGLSAPNSCISDEBQNH 974  
 Qy 1381 OGOSCIHRPCGKODSCPSLLLDHADVNCISIGELMKCATTCORGFALQASGQYTRPM 1440  
 Db 975 OGOSCIHRPCGKODSCPSLLLDHADVNCISIGELMKCATTCORGFALQASGQYTRPM 1034  
 Qy 1441 OKEILLTSSGHWDPQNSCLPVDGVPDPSPSLVNTANFSCSEGTFLKRCISICVPKALQ 1500  
 Db 1035 OKEILLTSSGHWDPQNSCLPVDGVPDPSPSLVNTANFSCSEGTFLKRCISICVPKALQ 1094  
 Qy 1501 GLSPMLTLEDGLWSLPEVYCKLECDAPPIILNANULLPHCLQDNHVDGTCCKYCKRGY 1560  
 Db 1095 GLSPMLTLEDGLWSLPEVYCKLECDAPPIILNANULLPHCLQDNHVDGTCCKYCKRGY 1154  
 Qy 1561 YVABSAEGVNRKLLKIOCLEGIEWOGSCIPVCEPPEVPEGYECTNGFSLDSQVL 1620  
 Db 1155 YVABSAEGVNRKLLKIOCLEGIEWOGSCIPVCEPPEVPEGYECTNGFSLDSQVL 1214  
 Qy 1621 NCNOREKLPILCTKGLWTOEFKLCENIOGCEPPPEPPLNSVEYKCEGYSIGAVCSPL 1680  
 Db 1215 NCNOREKLPILCTKGLWTOEFKLCENIOGCEPPPEPPLNSVEYKCEGYSIGAVCSPL 1274  
 Qy 1681 CVIPSPDPVMLPENTADLTLEHMEPVKQSVICTGRQOMHPVUHLCTIOGCEPQADG 1740  
 Db 1275 CVIPSPDPVMLPENTADLTLEHMEPVKQSVICTGRQOMHPVUHLCTIOGCEPQADG 1334  
 Qy 1741 WCDTINNRAICHYDGDCCSSTLSKVIIPFADCDLDECTCRDPAEENQ 1791  
 Db 1335 WCDTINNRAICHYDGDCCSSTLSKVIIPFADCDLDECTCRDPAEENQ 1385

## RESULT 5

US-09-983-025-25

Sequence 25 Application US/09983025

Publication No. US20030124529A1

GENERAL INFORMATION:

APPLICANT: OXVIG, Claus

TITLE OF INVENTION: PREGNANCY-ASSOCIATED PLASMA PROTEIN-A2 (PAPP-A2)

FILE REFERENCE: OXVIG-1A

CURRENT APPLICATION NUMBER: US/09/983, 025

PRIOR FILING DATE: 2000-10-20

PRIOR APPLICATION NUMBER: DK PA 2000 01571

NUMBER OF SEQ ID NOS: 25

SOFTWARE: PatentIn version 3.1

SEQ ID NO 25

LENGTH: 1627

TYPE: PR

ORGANISM: Homo sapiens

US-09-983-025-25

Query Match 39.7%; Score 3916.5; DB 11; Length 1627;

Best Local Similarity 45.8%; Pred. No. 1.5e-306;

Matches 718; Conservative 296; Mismatches 499; Indels 53; Gaps 25;

249 REAFTEFNSOVL P--IIYFSGRRERL-LRPREVLAIPREAFTEVAWKEGCONNPAT 305  
 80 REAGATNEBPSPRALYFSGRGQLAVLRADL--ELPRDAFTLQWLRAGEGCRSPAVI 137  
 306 AGVDNCSHTVSDGMLGIRSGDKRKDRARFPFSLCTDRVKKATLIIHSRYPQCTWT 365  
 138 TGLYDKCSYISRDGKMWVGHHTISDQNDKDRFYFSLKTDRAKVTTINAHSTYLRQWV 197  
 366 HVAATYDGRHVALYVDGTQVASSLDGSGPLNSPFMACSRSLIGDSSDEGHYFRHGLGT 425  
 198 YLAATYDQGFVKLYVNGAQVATSGEYQVGFSPLOCKCKVLMJGG--SALNHNRYGIEH 255  
 426 LVFSTALPQSHFGHSSQHSGBEADVLVTASPEPNTWVFPREKPRLEV--LQG 483  
 256 FSLMWKATQREILSDMETGHANTALPQLLQENWMDNYKAMSPKSGSSKVEFSNAG 315  
 484 FERPEILSPLOPLCGQTVCDNVELISQYNGWYPLRGEKVIYQVNNICDEGLNPIVS 543  
 316 FLUD---TSLERPLOGTLCDNTEVIASTNQLSFRPKVYKRVNVLVEDHKNTPTVT 371  
 544 BEQIRLOHEALNEAFSRYNISWQLSVHQNSTLRHVLVNCBPSKIGNDHCPCEBHP 603  
 372 RBQVDFQHQLAEAKQYNIWELDVLEVSNSLARRLLANCDISKIDENCDPECHNT 431  
 604 LTGYDGGDR-LQGRCYSMNRDGLCHVECNMMLNDPDDGCCDPQVADYKTCFDPSP 662  
 432 LTGHGCGCRHLRHPAFVKQKNGVCDMDCNTERFNGGECDELTINVTQTCDFPSP 491  
 663 KRAYMSYKELKALDNLSTHFLNIYFASVSREDLGAATWPMWDXAVTHLGGIVLSPAY 722  
 492 HRAVLDVNEKNIILKIDGSTHNLTFKASSEBELAGVATWPMKALMLHGGIVLNSPBY 551  
 723 GMPGHTDTIHEVGHVGLYHVFVKVSEKSCNDPCKETVPMETGDLCACTAPRPSKL 782  
 552 GMPGHTDTIHEVGHVGLYHVFVKVSEKSCNDPCKETVPMETGDLCACTAPRPSKL 611  
 783 CREPESTPTCGTPRPGAPFTNMYSTYDNCNDNTNPQVARMHCYIDLVLVYQWTESRK 842  
 612 CDPGNGDTCGHSFNTYNNFMSYADDDCTDSTPQVARMHCYIDLVLVYQWTESRK 671  
 843 PRPIPRPMVIGQTNKSLTIHMLPRISGVYDRAAGSLGACTEDGTFRQVYHTASSRV 902  
 672 PAVVALAPVGLHTTDSVTLMEPRPIDGHPFERELGASCHLCEGRILVQVYASNSMP 731  
 903 CDSGTYTPBEAVGPRVDOPCEPSLOANSPEVHLTHMMYVPCP-TEGCSILLFOHNV 961  
 732 CSPSGHMSPREAEGHPDVEQPCSVRTWSPNSAVNHTVPACPEPQCYLELEFLPL 791  
 962 QADTLTLMVT--SFPMSSQVLPDTLELLENKESVHLGPLDTFCDIPLTIKL-HYDGKVS 1018  
 792 VPESLITWTFVSTDMDSGAVNDIKLAVSGKISLGQNVACDVPILIRLMDVGEVY 851  
 1019 GVKVYTFEDERIEIDALLTSQSPSLCSGCRPVRYOVLDRPPASGLPVVYVTHSHKFPD 1078  
 852 GIGVYTLDEHLEIDAMLTSTADTPLCQCKPLKVKVVRPRLQMVASTL-HLNKFPD 910  
 1079 VERTPGMTOYOLAEGELGELASPRLNHIDHAPVCGDKVSEBRLGEECDPDLVSGDG 1138  
 911 MDNLGSGVYQWYVTTISGESESPSAVYTIHGGYCGDGI IQDQGEBCDDMKKINGG 970  
 1139 CSKVCLEEGFNCVGEPSLCAWTEGDCICEPFEKRTSIVCGIYTPKGYLDQWATRAS 1198  
 971 CSLFCRQEVSNFCIDBSRCYFHDGDCVCEFEKRTSIXKOGVYTPPGFLDQWASNASV 1030  
 1199 HEDKKKCPVSLVTEP-HSLICTSYHDLRNHRPLTGWPCVASENETQDRSEQPEGSL 1257  
 1031 HQD-QQCPGWIIGOPASQVCRTKVIDLSEGISQHAMVYCTISYPSQ----- 1078  
 1258 KKEDEWMLKCFNRPGEARAFILTTDGLVGEHQPTVLTLYTVDRGSHSGLTGLS 1317  
 1079 LAQTFMLRAVFSQPMVAAAVIYHLVTDGYTGQOKCTISVOLLDTKQSHDGLHVL 1138  
 318 COHNPILINVTHQVLEFHTTSTVILNFPSPVGI SAVALTSSRIGLSAPNSCISEDEG 1377

1139 CANNLILIPVHDLSDQPRYHQAIVRSFSSPLVAISGVALBSFNDPVTLSLSC-ORGET 1197  
 1378 QNHQOQSCITHRCPGKQDSCEPSLIDHADVNTCTSI---GGLMKCAITOCRGALASS 1433  
 1198 YSPAEQSCVHFACTKD-CPELAVENAS-LNCCSSDRYHG--AQCTVSCRTGYVLQIR 1252  
 1434 GQYIRPMQ--KEIILTCSSGHMDQVNSCLPYDCGVDPSPVLNVYANFSGSEGTKEFKCSI 1491  
 1253 DBELIKSQGPBYVYTTCTBGRKKNQVACEPDCSIDPHNQVYASFSCEGETFGSQCSF 1312  
 1492 SCVPPAKIQLGSLPWLTCLEEDLMSLPEYCYKLECDAPPIIANMLLPHCLQDNHVDGTI 1551  
 1313 QCRHPAQLKGNNSLITLCEMEDGLMSPEALCELMCLAPPVPVADIQTARCBENKRVGSF 1372  
 1552 CKYCEKPGYVVAESABGRKRNLLKICLEGIGIEQSGCIPVYCEPPEVEGMECTING 1611  
 1373 CKYKCPGYNHVGSSR-KSKKRAFKTQCTQDQSGMBGACVPTCDPPRPKFGYLGCTNG 1431  
 1612 FSLDSQCVLNC-----NQEREKLPILCTKEGLMTQEFKLCENLQGECPRPSELNS-VEX 1665  
 1432 FQNSSECRKCEBDSASQGLGSNVITHCRKDTWNGSFHVCQEMQGC-SVPVELNSNKL 1490  
 1666 KCBQYIGAVGSPCLVTPSPDPWLPENITADTLEHMMEPVKQSIYCTGRKQHPV 1725  
 1491 QCPDGYAIGSECATSCLDHNSSEIILPMNVYRDIPLHMLNPRVERVUCTAGLKMYRPA 1550  
 1726 LVNJCIOSCPFOADGMDTINRAYCHYDGGCCSSTLSKVIYFPAADCDLD-BCTGRD 1784  
 1551 LHCIVGCEPFGMDNYCDALNNRAFCNYDGGCCTSTVTKKVTTPRMSQDQSGACGD 1610  
 1785 PKAEN 1790  
 1611 PQQEH 1616

RESULT 6  
 US-10-295-027-663  
 Sequence 663, Application US/10295027  
 Publication No. US2003023350A1  
 GENERAL INFORMATION:  
 APPLICANT: Afar, Daniel  
 APPLICANT: Aziz, Nacasha  
 APPLICANT: Ginsberg, Wendy M.  
 APPLICANT: Glash, Kurt C.  
 APPLICANT: Glynn, Richard  
 APPLICANT: Hevezl, Peter A.  
 APPLICANT: Mack, David H.  
 APPLICANT: Murray, Richard  
 APPLICANT: Watson, Susan R.  
 APPLICANT: Eos Biotechnology, Inc.  
 TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and  
 FILE REFERENCE: 018501-012500US  
 CURRENT APPLICATION NUMBER: US/10/295, 027  
 CURRENT FILING DATE: 2002-11-13  
 PRIOR APPLICATION NUMBER: US 09/663, 733  
 PRIOR FILING DATE: 2000-09-15  
 PRIOR APPLICATION NUMBER: US 60/350, 666  
 PRIOR FILING DATE: 2001-11-13  
 PRIOR APPLICATION NUMBER: US 60/335, 394  
 PRIOR FILING DATE: 2001-11-15  
 PRIOR APPLICATION NUMBER: US 60/332, 464  
 PRIOR FILING DATE: 2001-11-21  
 PRIOR APPLICATION NUMBER: US 60/334, 393  
 PRIOR FILING DATE: 2001-11-29  
 PRIOR APPLICATION NUMBER: US 60/340, 376  
 PRIOR FILING DATE: 2001-12-14  
 PRIOR APPLICATION NUMBER: US 60/347, 211  
 PRIOR FILING DATE: 2002-01-08  
 PRIOR APPLICATION NUMBER: US 60/347, 349  
 PRIOR FILING DATE: 2002-01-10  
 PRIOR APPLICATION NUMBER: US 60/355, 250

PRIOR FILING DATE: 2002-02-08  
PRIOR APPLICATION NUMBER: US 60/356,714  
PRIOR FILING DATE: 2002-02-13  
Remaining Prior Application data removed - See File Wrapper or PAM.  
NUMBER OF SEQ ID NOS: 1386  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 663  
LENGTH: 1627  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-295-027-663

Query Match 39.7%; Score 3916.5; DB 12; Length 1627;  
Best Local Similarity 45.8%; Pred. No. 1.5e-306;  
Matches 718; Conservative 296; Mismatches 499; Indels 53; Gaps 25;

QY 249 REAFTRISQGLP--LIFESGRERL-LARPEVLAETPREAFYEAAMVKEGGONPAIT 305  
DB REARGATEEPSPSRALYFSGRGEQLRLVRLADL-ELPRDAFTLVWLRAEGQSRPAVI 137  
QY 306 AGVFNDNSHTVSDKGMALGIRSGDKDKRDARFPFSLCTDRVKKATILISHRYOQPTWT 365  
DB 138 TGIYDKCSYISRBGRVVGIIHTISDQNDKPRYFSLKTRBRAQVTTINHRSTYLPQNV 197  
QY 366 HVAATYDGRMALYDGTQVASSLDGSPINSPFMASSRLLEGDSSEDEGHYFRGHGT 425  
DB 198 YLAATYDQGMKLYVNGAQVATSGEQVGFSPITQCKVLMGG--SALNNHYRGYIEH 255  
QY 426 LVWATSLPQSHOHSSGSGEAEATDLYLTASFEEVNTVWVPRDEKYPRLAEV--LOG 483  
DB 256 FSLMKVARTQREILSDMETGATHTALPQLLOEMWVVKIAMSFMKGGSPKYEFSNAG 315  
QY 484 FEDEPEILSLOPLCGQVCDNVELISOYNGWPLRGEVIRIOVNICDDGLNPIS 543  
DB 316 FLID---LSLEPPLCGQLCDNTEVASTYQSSFPQPKYVRIRVNLVEDDHKNFTVT 371  
QY 544 EEOIRLOHEALNAFNRYNISWOLSVHQVNSTLRHVLVNCESPKIGNDHCEPCEHP 603  
DB 372 REGVDPQHQLAFAKQYNISWELDLVENSISLRRLILANDDISIGBENCDEPNHT 431  
QY 604 LTGYDGGDCR-LOGRCYSNNRBDGLCHVECNMNLNDFDDCCDQVADYRKTCEPDPSP 662  
DB 432 LTGHGGDCCHLHRAFKQHNHGVCDMDCYERFNDGECCECPETITVYQTCFDPDSP 491  
QY 663 KRAYMSYKELKALQNLNSTHPLNYFASVREDIAGAATWPMDDATHTHIGIYLSAPY 722  
DB 492 HRAYLDNBLKNLTKLDGSHLINFPAKSSBEBELGVALTWPMDEKALMHGGLVNSFY 551  
QY 723 GMPGHTDTMHEVGNVLGLYHVFEGVSESSCNDPCKETVPSMETGDLCACTAPTPKSEL 782  
DB 552 GMPGHTDTMHEVGNVLGLYHVFEGVSESSCNDPCKETVPSMETGDLCACTAPTPKSEL 782  
QY 783 CREDEPTSDTCGTRPRGAPPTNYMSTTDNCTDNFTPNQVARNHGYLDLVYQOMTESRK 842  
DB 612 CGPDPGENDTCGFSFNTYNNFMSYADDDCTDSFTRPNQVARNHGYLDLVYQOMTESRK 842  
QY 843 PIRIPRPMVIGQNKSLTIHMLPRISGVVYDRASSGLSCAGCTEDGFRQVHTASRRV 902  
DB 672 PAPALAPQVGLHTDSTVLEWPFRLDGHFERELGASCHLCLEGRLLVYASNASSBMP 731  
QY 903 CDSGGVTPREAVAPRPVDPDCEPSLQAMSPEVLYHMMNTVPCP-TEGCSLELTFPHV 961  
DB 732 CDSGSHSPREABEHPVBOCKSSVATGSPNSVNNHTVPRACREPOGCTYELFEFYPL 791  
QY 962 QADTLTLMVT--SFFMESOVLPFTEILLNKESVHLGRLDTFCIDIPYTKL-HVDGKVS 1018  
DB 792 VPESLITWVTFVSDMDSSGAVNDIKILAVSGKISISLQGVNFCVDTLIRLMVGEVY 851  
QY 1019 GVAVYTPDEDELEIDAALTGQPHSLCSGCRPVAYOVLDRPFAASGLPVVYVTHSHRFTD 1078  
DB 852 GIDQYTLDEHLEIDAALTGQPHSLCSGCRPVAYOVLDRPFAASGLPVVYVTHSHRFTD 1078  
QY 1079 VEVTPGOMYOYVLAEGELGEASPLNHTHGAHYCGDGKVSERLGECCDDGLVSGDG 1138

DB 911 MDNLASVYQWVYITISGTESESPAVTYIHGRGYCGDGIIOKDOGEOCCDDMKINGDG 970  
QY 1139 CSKVCLESGFNVGVSPSICMYWEGDICEPPEPKRTSYIDCGIYTKGVLDMATAYSS 1198  
DB 971 CSLFROEVSFNCTIDPSRCYFPDGDGCEBEFOKTSIDCGVYTOGFLDOASNAVS 1030  
QY 1199 HEDKKCPVSLYTGEB-HSLICTSYHAPDLPNHRPLTWGPFVASENEODDRSEOGESL 1257  
DB 1031 HDQ-QCCPQWVITIGPAAQVCEKTKYIDLSBGSISQAMWYCTISYSQ----- 1078  
QY 1258 KKEDEVMLKVCENRGEBAIFLELTGDLVGEHOQPTVYLYLDVRSNHSGLTYGLS 1317  
DB 1079 LQOTFTMLRAYFQOPVAAVYIHLVTDGYDQKQETISVQLDITKQSHDLGHLVLS 1138  
QY 1318 CQHNPLIINTVHTQNLPHHTTSVLLNFSSPRGISAVLRSRIGLASPNSCISEDEG 1377  
DB 1139 CRNNPLIIPVVDLSOPFHSQAVRVSFSSPLAIGVALRSFDFNPVTLSSC-ORGET 1197  
QY 1378 QNHQGSCTHRPCGKODSCPSLLDHDADVNCSTSI----GPGIMKCAITCORFALQASS 1433  
DB 1198 YSPABQSCVHFACETKD-CPELAVENAS-LNGSSDRYHG--AQCTVSCRTGYVLQIR 1252  
QY 1434 GQYIRPMQ--KEILLTSSGHWQDQVNSCLPVDGVPDPSLVNVAHFSCSEGTFLKRCST 1491  
DB 1253 DBELIKSQGSPSVYVCTEGKMNKQVACEPVDCSIIDHHQVYVASFSCPEGTFGSGCSF 1312  
QY 1492 SCVPRAKQLGSLSWLTCLEDEGLMSLPEVYCKLECDAPRIILNANLLPHLQDNHVGTT 1551  
DB 1313 QCHPAPQLKGNNSLLTCMBEGILMSFPBALCELMCLAPPVPVNDLDTARREKHKVGSF 1372  
QY 1552 CKYCKRGVYVVASAGKVRNKLKIOCLEGGIWEQSSCIPVCEPPPEFEGVECTNG 1611  
DB 1373 CKYCKRGVYVVASAGKVRNKLKIOCLEGGIWEQSSCIPVCEPPPEFEGVECTNG 1611  
QY 1612 FSLDQCVLNC-----NOEREXKLPICTKGLWTOBFKLCENLOGSCPPPESELS-VEX 1665  
DB 1432 FQNSCRKICEDSDASQGLGSNVHCRDGTNGSFHVQOEQOQC-SVFNELNSYLKL 1490  
QY 1666 KCEQYIGAVGCPPLCYTPPSDPMLEPENTADTLEHMEPVVQSYVCTGRQMHDPV 1725  
DB 1491 QCPDGYAIGSECTSCIDHNSESIIIPMNYVADIPHMVPTVERVCTAGAKWYHPA 1550  
QY 1726 LVHICQSCPEFQADGCDTINRAYCHYDQGCCSSSTLSKQVIFPAADCDL-ECTCRD 1784  
DB 1551 LHVCHVGCPEFMDNYCDALNRAFCNYDGGDCTSTVTKKVTFFPMSCDLDGDCACRD 1610  
QY 1785 PKAEN 1790  
DB 1611 PQAOEH 1616

RESULT 7  
US-09-864-761-34265  
Sequence 34265, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wenhang  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: Aeomica-X-1  
CURRENT APPLICATION NUMBER: US/09/864, 761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263,6



```

; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 34265
; LENGTH: 192
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL031734.9
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 44
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.1
; OTHER INFORMATION: EST HUMAN HIT: AUI40701.1, EVALUE 2.00e-53
; OTHER INFORMATION: SWISSPROT HIT: P07207, EVALUE 3.00e-04
US-09-864-761-34265
```

```

Query Match 11.0%; Score 1086; DB 9; Length 192;
Best Local Similarity 100.0%; Pred. No. 2.5e-79;
Matches 192; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

Qy 472 DEKTPRLVLOGFPEPEPILSPLOPPLCGQTVCDNVELISQYNGWYPLRGEKVIKYQVN 531
Db 1 DEKTPRLVLOGFPEPEPILSPLOPPLCGQTVCDNVELISQYNGWYPLRGEKVIKYQVN 60
Qy 532 ICDEGLNPIVSEQIRIQHEALNEAFSRVNI SWQLSVHYVHNSTLRHRVVLVNCPEPSKI 591
Db 61 ICDEGLNPIVSEQIRIQHEALNEAFSRVNI SWQLSVHYVHNSTLRHRVVLVNCPEPSKI 120
Qy 592 GNDHCDPECEHPLTGYDGGDCRLQRCYSNMRDGLCHVECNMMLNDPDDCCDDPOYAD 651
Db 121 GNDHCDPECEHPLTGYDGGDCRLQRCYSNMRDGLCHVECNMMLNDPDDCCDDPOYAD 180
Qy 652 VRKTCFDPDSPK 663
Db 181 VRKTCFDPDSPK 192
```

RESULT 8

```

US-09-864-761-34264
; Sequence 34264, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aemica-X-1
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 34264
; LENGTH: 70
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL031734.9
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 44
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.1
; OTHER INFORMATION: SWISSPROT HIT: P28977, EVALUE 1.60e+00
; OTHER INFORMATION: EST HUMAN HIT: BF366974.1, EVALUE 4.00e-36
US-09-864-761-34264

Query Match 3.9%; Score 383; DB 9; Length 70;
Best Local Similarity 100.0%; Pred. No. 3.1e-23;
```





US-10-028-248A-47

Query Match 3.5%; Score 346.5; DB 12; Length 3567;  
 Best Local Similarity 19.3%; Pred. No. 1.5e-17;  
 Matches 344; Conservative 191; Mismatches 623; Indels 629; Gaps 99;

286 AFTVEAWKPEGGONNPAIIAGVF---DNCSHTVSDKGMALGIRSGKCKRDRARFFS 341  
 1449 AITCAFMMKSDVINYGPISYALBEDDKDNTSLTDNNGWLVY-NGEK- 1497  
 342 LCTDRVKKATLISHSRYPQGTWTHAATY--DGRHMAIYD-----GTVAASLIDQ 391  
 1498 -----ITNCPNVNDIMTHIATITWTSTGAMRVYINGELSDGTLSIGKLIPIG 1546  
 392 SGPLNSPRMASCRLILGGDSSEDEGHYR-----GHGLTVFVSTAL--PQSHQHSQH 444  
 1547 GG-----ALVIGQBODKKGEENPABSFVGSISQNLMDVYLSPOOVKLMS-- 1593  
 445 SSGEEATDVLVTASFEPVNTWVFPDEKYPRLV-----LQGFEEPEPILS 492  
 1594 SCPELSRGVLA-----WDFLSGITGKVKVNDSSMFCSDCPSELBSVHLRPAAS 1644  
 493 PLQPLPGQVY--CD-NVELIS--QY--NGTW--PLRGEKVIYQVNVICDEGLNP 540  
 1645 GNRKP--GSKVSLFCDFGFMVGNPVQYCLNQGQWTOPLPCHERIR-----C--GLPP 1693  
 541 IV-----SEQIRIQHEALNEAFSRV-----NISMQSLVQVNHSTLRHRYL 583  
 1694 ALENGFYSAEDFHAGSTVYTCSTGYLLGDSRMFCTDNGSMN----- 1736  
 584 VNCEPSKIGNDCHDCEPEHPLTGYDGDGDCRLQGRCYWNRDRGLCHVECNMLNDFDDGD 643  
 1737 -GISPSCIDVDECAV-----GSDCSHASCALNN--GAVYGCNPRPYG-DQKN 1781  
 644 CDDPOVADVKTCPDPDPSPKRAYSVKELKALQNLNTHFLNTYFASVREDLAGAA-- 700  
 1782 CAEP--VKCKAPENBENGHS--GEIYTVGTA--VFSSCEGHEHLVGVSTIT 1827  
 701 ---TMPMDK---DAVTHLGGIVLSPAYGMPGHTDMIHVGVGLVYHVFQVS-- 749  
 1828 CLEFGEWRLRPPSCAIS--CGVPVPENGVDSGAFYVGSKV-----VYRCDKYTLISG 1880  
 750 -ERESCDNPKCEVPSMETGDLCADTAPTKSELCREPEPTSD---TCGTFRFGAPFT 804  
 1881 DEESAC-----LASGS-WSHSSPVCELVKCSQPEDINNGKYLISGLT----- 1921  
 805 NYMSYTDNDCTDNTFPNOVARNHCYLDLYOQWTESRKPTPIPIPPWYIGQ--TNKSLTI 862  
 1922 -YLSIASYSCENGYSLQGPSLLECTASGSMWRAPPSQCVASGCEPPIYKDAVITGSNFTF 1980  
 863 HMLPPIISGVVYDRASGSLCGACTEDGTFRQYVHTASSRVCDSSGYWTPBE--AVG-- 916  
 1981 -----GNTVAYTCKEG--YTLAGPDITVCOANKMWSNMHQCCLAVASCD 2021  
 917 -PPVDQCPCEPSLQAMSPVHLYHMMNTVPCPTBGCSLELLFOHPVQADTLTLWMTSFFM 975  
 2022 EPPVNDHA-----SPET-----AHLRBDT-----APYYC 2046  
 976 ESSOVLFTETILLENKESVHLGP-----LDYFCDIPLTILKLVHDKVSGVXY- 1022  
 2047 ADGYSLADNSOLICNAQGNWVPBAQAVPRCAHCEKRPVSYSILLESVAKAPRAGSV 2106  
 1023 -----YTFDEREIDAALLITSPHSLPGSGCPRYVYOLRDP--FASGLP----- 1066  
 2107 VSPKMEGFVNTSAKIECLRGEMSPPLSVQCIPIVR--CGEPPIANGPSTGNTSYF 2163  
 1067 ---VVVTHSHRKFTDVEVTPGQVY--QYQVLAEGSELGEASPPLNHI--HGAPYCGDGX 1119  
 2164 GAVVAVYSGHKGf-----YIKGKKSTCEATGQMSKPTPTCHPVPSCNRPVENGf 2213  
 1120 VSERLGE-----ECDGDGLVSGD-----GCSK----- 1141  
 2214 LEHTTGRFESSEARPCQNPVYKAAGSPVFCQANRHMHSADAPLCTPLNCGKPPPIQNGF 2273

1142 -----VCELEGFNCVGPESLCTMYWEGDICEPERKTSIVDCGIYTPKGY 1187  
 2274 LKGESEFVSGKQVFC--NEGIELVGDNSWTQCKSGMSKRP---SPKCVPTKCAEPPL 2328  
 1188 LDQWATRAYSHEDKKKCPVSLVT--GEPSHICTSYHNDLPNHRPLTWMP-----C 1238  
 2229 ENQVLKELASB-----YGVMTISCKEGHALQGPVLKCLPSGQ--WNGSPICKMVL 2380  
 1239 VASENETQDDRSQEPESLKKEDBEVWLKVFENRGEARAFI-----FLTTDGLVREHQ 1293  
 2381 -----PSPPL-----IPFGVPASSGALHFGSYKYLCLVDGF--LRG 2415  
 1294 QPVTLYLTVRGSNHLITGYLSCQHNPLIIN-VTHNQVLPHNTSYVLNFSFSPVGI 1352  
 2416 SPIT-LQADSTWSSPLRECVPCBPPELINCIIHVQGLAVISTLYLCKRGFELVG- 2473  
 1353 SAVAL--RTSSRIG--LSAPNSCISEDEGQNHQ-----GQ----- 1383  
 2474 NATLCEGNGQWLQKRMCKPIECPEPKIINGQFSSVSFOYQOTIYFCDRGFLREGPK 2533  
 1384 --SCIHPCGKQDSCPSLLDHDVNVCTSIGP--GLMCA-----ITQORFAL 1429  
 2534 SLTCLF--TGDMWMDP--PSCDAIHCDPQPLENGFVVGADYRGAMITYSQFPQFQV 2587  
 1430 QASGQYIRPMQKEIILTCSSGHW--DONVSCLPVDCGVPD----- 1468  
 2588 LGHAMQ-----TCESGKSSSPTCVPIIDCLPHRIDGCTKRDQGHDOE 2636  
 1469 -----PSLVYANFSCSEBGTFL--KRCSISCVPAKQGLSPWLT 1507  
 2637 DDMVEVRYLAHPQLBATAVALAVENTKESPASHASHFLYGTWVSYSCEPYELLGI--PVL 2695  
 1508 CLEGLMSLPEVYC-KLECAPPIILNANILPHICODNHDVGTICKCEKPGYVAESA 1566  
 2696 QQEDGTWNGTAPSCISTIECDLPVAPENGLF--HTQYTT--KMSAAYSCPKHILIEGSH 2750  
 1567 EKVYRNKLKIQCLEGGIMEGGS--CIPVCEBPPEPVFEG-----MYECTN 1610  
 2751 -----LRL-CLQNKQW--SGTYPRCEALISCSKPNLNMNGSIKGDYSLVLYYICDS 2800  
 1611 GFSLDSCVLCNQNQERKELPLCTKEGLWTOEFKLCENLQCECPPEPSELN----- 1661  
 2801 GYILNKKRKTQENRD-----MDGHEBMC--IPVDCSPVPPTNGRVKGEY 2847  
 1662 ---SVEIKCEQGY-----GIAVCGPL--CVLPSPVWLYPEVITADTL 1700  
 2848 FOKEITVSCREGFLLEGARSRICTLNGSWGATPSCMPVRCBPAPQVP-----NGVADG 2902  
 1701 E-----HMMEPVAYQS--IVCTGRQWHPDPVLVHCIOGCEP 1735  
 2903 DYCFKKEVAFHCLGELYVLOQAPRLTQCSNGTWDAE-----VPVCKP 2943

RESULT 11  
 US-09-911-842-4  
 ; Sequence 4, Application US/09911842  
 ; Patent No. US20020151483A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Welcher, Andrew A.  
 ; APPLICANT: Elliott, Gary S.  
 ; TITLE OF INVENTION: C3B/C4B COMPLEMENT RECEPTOR-LIKE MOLECULES AND USES THEREOF  
 ; FILE REFERENCE: 0101737392  
 ; CURRENT FILING DATE: 2001-07-24  
 ; PRIOR APPLICATION NUMBER: US 60/222,438  
 ; PRIOR FILING DATE: 2000-08-01  
 ; NUMBER OF SEQ ID NOS: 5  
 ; SOFTWARE: PatentIn version 3.0  
 ; SEQ ID NO 4  
 ; LENGTH: 3594  
 ; TYPE: PRT  
 ; ORGANISM: Mus musculus

FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (1757)..  
 ; OTHER INFORMATION: Xaa = any or unknown amino acid  
 US-09-911-842-4

Query Match 3.4%; Score 336.5; DB 10; Length 3594;  
 Best Local Similarity 19.2%; Pred. No. 9.7e-17;  
 Matches 343; Conservative 189; Mismatches 626; Indels 629; Gaps 99;

QY 286 AFYEAAYKPEGGONPAIYAGV---DNCSTHVSDEKGMALGIRSGDKKRRARFFFS 341  
 DB 1476 AVTCAFMKSSDVINYGPISYALEDDKNTFLITDYGWVLYV-NGKEK----- 1524  
 QY 342 LCTBRVKATILISHSRQPGTWHVATVD--GRHMLAYVD-----GTQVASSLDQ 391  
 DB 1525 -----ITNCPVNDGIMHIAITMTSIGAMFVYIDGELSDGCTLSIGKALPG 1573  
 QY 392 SGPLNSPMACSRLLSGDSSBDGNYFR-----GHIGTLVFWSTAL-POSHFOHSSQH 444  
 DB 1574 GG-----ALVIGQEDKKGGEGFNPAESFVGSISQNLMDYVLSPOQVLLAS-- 1620  
 QY 445 SSGSEBATDVLVTSFEPVNTWVFPFDEKYPRLFY-----LQGEPEPELIS 492  
 DB 1621 SCPELISRGVLA-----WDFLSGIGTKVAVDSSSMFCSDPELSEGVHLRPA 1671  
 QY 493 PLQPLQGVY---CD-NVELIS---QY---NGVW--PLRGEKYRYGVNVICDEGLNP 540  
 DB 1672 GNRKP--GSKVSLFCDFGQFQVGNFVQYCLNOGQTOPLPCHERLR-----C--GLP 1720  
 QY 541 IV-----SEQIRLOHEALNEAFSRY-----NISWOLSVHVNSTLRHRYVL 583  
 DB 1721 ALENGFIYAEDFHAGISTYTOCTSGYVLLGDSRMFCXNGSMN----- 1763  
 QY 584 VNCEBSKIGNDHCPRECHPLTGTIDGDCRLQGRCYSNRRDGLCHEVCNNMLNDFDDGD 643  
 DB 1764 -GISPSCLDVDECAV-----GSDCSEHASCINTN---GSYYCSGNEPPTYG-DGKN 1808  
 QY 644 CCDDQVADVNRKTCFDPDSPKRAYMVKELKEALQNSTHFLINTYPAASVRDLGAA--- 700  
 DB 1809 CAEP-----VKCKAPENPERGRSS---GETYVYGA---VTFSCDEGHELVGVSTTT 1854  
 QY 701 ---TWPMDK-----DAVTHLGGIYVLSPAYVGMFGHTDMIHVGHLVLYHFVKS--- 749  
 DB 1855 CLFTEGMRLRPLSCAIS--CGVPPVBNGVDSGAFYVGSV-----VYKCDKYTISG 1907  
 QY 750 -ERESCDNPCKEYVPSMETGDLCACTAPTPKSELCREPEPTSD---TCGFTRPGABFT 804  
 DB 1908 DEESAC-----LASGS-WSHSSPVCGLVKCSQPEDINNGKYLISGLT----- 1948  
 QY 805 NYMSTTDNCTDNFTPNQVARNHCYLDLYVOQWTSRKPTPIPIPMYIGQ--TNKSLTI 862  
 DB 1949 -YLSIASYSCENGYSLQGPSLLECTASGWDAPSCQVSCGEPPIYKDAVITGNSNTP 2007  
 QY 863 HMLPISGVVYDRAAGSLGACTEDGTFRQYVHTASSRRVCSGYMTPER---AVG-- 916  
 DB 2008 -----GNTVAYTCKEG---YTLAPDIIICQANGKRNSSNHQCLAVGCD 2048  
 QY 917 -PPVDVQCEPSLQAMSPPEVHLVHNMVTPCPTEGCSLELLFQHPVADTLTLMTSPFM 975  
 DB 2049 EPPNVDA-----SPET-----AHLRFGDT-----AFYYC 2073  
 QY 976 ESSQVLFTEILLKESVHLGP-----LDTRCDIPLTIKLVADGKVSQVK--- 1022  
 DB 2074 ADGYSLAONSOLICNAQGNWVPACQAVPRCIANHCERPSVYSILSSVAKPAAGSV 2133  
 QY 1023 -----YTFDERLEIDALLTSQHPSLGSGCPVRVQVLRDP--FASGP----- 1066  
 DB 2134 VSFKMBEFLVNTSAKIECLRGEMSPSLSVOCIPVR---CGEPFLSANGPISGTYNSF 2190  
 QY 1067 --VVVTHSHRFTDVEVTPGQNY---QYQVLAEGAGELGASPPPLNHI--HGAPYCGDK 1119  
 DB 2191 GAVVAVSGCHKP-----YIKGEKSTCEATGQMSKPRPTCHPVSQCNBPVKEG 2240

QY 1120 USERIGE-----ECDDGDLVSGD-----GCSK----- 1141  
 DB 2241 LEHTTGTFPSEARFQGNPCGYKAAAGSPVFOQANRHHNSDALPSTPLNCGKPPPIONGF 2300  
 QY 1142 -----VCELEBGFNCVGPESLCTMYEGDICEPPEKRTSIYDCGITYPKY 1187  
 DB 2301 LKGSFPEVSGKVQVVC--NEGVELYGDNSWTQCKSGKMSKKP--SPKCVPTKCAEPRL 2355  
 QY 1188 LDQWATAYASHHEKCKCPVSYT--GEPHSILCTSYHNDLPYHRLTGWFP-----C 1238  
 DB 2356 ENQVLKELASE-----GVMTISCKEHALQGPVLKCLPSGQ--WNSFPICRMYLC 2407  
 QY 1239 VASENETODRSEQPEGSLLKKEDEVMKVCFNRRGEARAFI-----FLTDLGLVPGBHQ 1293  
 DB 2408 -----PSRPL-----IPGVASSGALHFGSTVXKLCVDFPF--LRG 2442  
 QY 1294 QPVTLYLTVDRGSNHSLGTYGLSCQHNPLIIN-VTHQNVLFHNTTSVLNFSPPRYGI 1352  
 DB 2443 SPRT-LCADSTWSSPLPECVAVPCQPQPEELINGIIHVQGLAYSTLYTCKPGEFLVG- 2500  
 QY 1353 SAVAL--RTSSRG--LSAPNSCISEEGQNHQ-----GQ----- 1383  
 DB 2501 NATLCEGNOQWLGKPMCKPTECEPEKELNGQFSSVSFOYGTITYPCDRGFRLEBPK 2560  
 QY 1384 --SCIHPPCGKODSCPBLLDHADVNCTSIGP--GLMKCA-----ITCORGFAL 1429  
 DB 2561 SLTCLF--TGDMWMDP--PSCDAIHCSDPQRIENGFVEGADRYGMIIYSCFPGRQV 2614  
 QY 1430 QASSGQYIRPMQKILLTCSGGHW-DQNVSCLPVDCVDP----- 1468  
 DB 2615 LGHAMQ-----TCBSGWSSSPTCPVPCDGLPHI1DGDCTKVADGQGHFPOE 2663  
 QY 1469 -----PSLVYANFSCSEGRKPL--XKCSIQVPRALQSLPPLT 1507  
 DB 2664 DDMNEVRYLAHPHLEATATALENTKESPAASHAPLYGTWVSSCEGYELLSI-PVLI 2722  
 QY 1508 CLEDELMSLPVYC-KLECDAPRIILNANLLPCLDNDHNVGTICKCKEKGYYVASEA 1566  
 DB 2723 QOEDGTWNGTRPSCISIECDLPVAPBNGFL--HFTQTT--MGSAAQVSCRGHILBESH 2777  
 QY 1567 EGVYRNKYLKQCLEGGIWEQGS--CIPVCEBPFPVFEQ-----MYECTN 1610  
 DB 2778 -----LRH-CLQNNQW--SGTVPRCEALISCKRNPLMNSIKGDVSYLVLYECDS 2827  
 QY 1611 GFSLDSQVLNCGNDEKELPLTCKEGLWTOEFKCEYLQSECPPEPSELN----- 1661  
 DB 2828 GYILNGSKKRTCOENRD-----WDGHEPNC--IPVDCGSPVPTGRVKGEEYT 2874  
 QY 1662 --SVEYKCEQY-----GIGAVCSPL-CVIRPSDPVMLPENITADTL 1700  
 DB 2875 FOKEITYSCREGFLLEGARSRICLTNWSGATPSCNRPVRCAPAPQV--NGVADGL 2929  
 QY 1701 E-----HMEPVRVQVS--IVCTGRQWHPDPVLYHICQSCBP 1735  
 DB 2930 DYGFKEVAFHCLGEGVYLQCAPRLTCQNGTWDAE-----VPYCKP 2970

RESULT 12  
 ; US-10-150-821-4  
 ; Sequence 4, Application US/10150821  
 ; Publication No. US20020192758A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Welcher, Andrew A.  
 ; APPLICANT: Eliott, Gary S.  
 ; TITLE OF INVENTION: C3B/C4B COMPLEMENT RECEPTOR-LIKE MOLECULES AND USES  
 ; TITLE OF INVENTION: THEREOF  
 ; FILE REFERENCE: 01017/37592  
 ; CURRENT APPLICATION NUMBER: US/10/150,821  
 ; CURRENT FILING DATE: 2002-05-16  
 ; PRIOR APPLICATION NUMBER: US/09/911,842  
 ; PRIOR FILING DATE: 2001-07-24  
 ; PRIOR APPLICATION NUMBER: US 60/222,438

PRIOR FILLING DATE: 2000-08-01  
 NUMBER OF SEQ ID NOS: 5  
 SOFTWARE: PatentIn version 3.10  
 SEQ ID NO 4  
 LENGTH: 3594  
 TYPE: PRT  
 ORGANISM: Mus musculus  
 FEATURE:  
 NAME/KEY: misc feature  
 LOCATION: (1757)..  
 OTHER INFORMATION: Xaa = any or unknown amino acid  
 US-10-150-821-4

Query Match 3.4%; Score 316.5; DB 14; Length 3594;  
 Best Local Similarity 19.2%; Pred. No. 9.7e-17;  
 Matches: 343; Conservative 169; Mismatches 626; Indels 629; Gaps 99;

QY 286 AFTVEAMVKEPGGONPAIIAGVF---DNCSTVSDKGMALGIRSGDKKRDARFFFS 341  
 DB 1476 AVTCAFMKSSDVINGVPISYALEDDKDNFTLLTDYNGVLYV-NGREK----- 1524  
 QY 342 LCTRVKATILISHSRQPGTWHVATYD--GRHMLLYD-----GTQVASSLDQ 391  
 DB 1525 -----ITNCPVNDGIMHIALITWTSIGAMRWYIDGLSDGTGLSIGKALPG 1573  
 QY 392 SGPLNSPFMASCRIILGDSSESDGHYR-----GHGLTVFWSTAL-POSHFOHSGH 444  
 DB 1574 GG-----ALVIGQEDKKGEKGNPAESFVGSISQALNDYLSPOQVILLAS-- 1620  
 QY 445 SSGSEATDLVLTASFEPVNTWVPEFDEKYPRLV-----LQGFEPPELIS 492  
 DB 1621 SCPELSRGNVLA-----WPDFLSGITGKVXVDSSMFCSDPELSGVHLRPAAS 1671  
 QY 493 PLOGPLCGQY---CD-NVELIS---QY---NGYV--PLRGEKIRYVVAICDDEGLN 540  
 DB 1672 GNRKP--GSKVSLPCDPGFQVGNPVOYCLNQGWTOPLPCEKIR-----C--GLRP 1720  
 QY 541 IV-----SEQIRLOHEALNEAFSRY-----NISMQLSVHQVHNSLIRHVVL 583  
 DB 1721 ALBNGFYSADPFHAGSYTYTQCTSGYLLIGSRMFCXNGSN----- 1763  
 QY 584 VNCEPSKIGNDHCDPECHPLTGYDGDCLQGRCYSWNRDGLCHVECNMMLDPDGD 643  
 DB 1764 -GISPCCLDVDECAV-----GSDCSEHASCLNTN--GSVYCSGPNPYTG-DQKN 1808  
 QY 644 CCDDQVADVKTCTDPPSPKRAYMSVKELKRALQNLNTHPLANTYPAASVREDLAGAA-- 700  
 DB 1809 CAEP-----VKCPAPENPENGSRG--GEITYVGT-----VTFSCDEGHELVGSYIT 1854  
 QY 701 ---TWPMDK-----DAVTHLGIIVLSPAYCGPHGTHMIEHVGVLGLYHFKGVS-- 749  
 DB 1855 CLEFGEHMDRLRPSCEALIS--CGVPVPENGVDGSAFTYGSKV-----VYCDKGYITLSG 1907  
 QY 750 -ERESCDNPCKETVPSMETGDLCADTAPTPKSELCREPEPTSD-----TCGTFRPGABFT 804  
 DB 1908 DEBSAC-----LASGS-WGSHSPVCGLVKCSQPEIDNNGKYLISGLT----- 1948  
 QY 805 NYMYTNDNCTDNTFPNOVAMHCLLDLYVOQWTESKRPPIPIPPMWYGO--TNKSLITI 862  
 DB 1949 -YLSIASYSCENGYSLOGPSLLECTASGSMWRAPSSCOLVSCGEPIYKDAVITGSNTP 2007  
 QY 863 HMLPPISSGVVDRASGSLCGACTEDGTFRQYVHTAASSRVDSSGYWTFEE---AVG-- 916  
 DB 2008 -----GNTVAATYCKEG--YTLAEPDTTICQANGKMSNHQCLAAVSCD 2048  
 QY 917 -PPVDQPCPSPLQAWSPPEVHLVHMNMVTPCPTGCSLELLFQHPVQADTLTLWVTSFFM 975  
 DB 2049 EPPVVDNA-----SPET-----AARLFEDT-----AFYYC 2073  
 QY 976 ESSQVLFPTLELLNKESVHLGP-----LDTFCDIPLTIKLAHDKGSGVKY----- 1022  
 DB 2074 ADGYSLADNSQLICNAQGNWVPAPGAQAVPRCIAHFCEKPSVSVSILSEVSKAKFAAGSV 2133

QY 1023 -----YTFDERIEIDAAALLTSQPSPLCSGCRPVRYOVLDRPP-FASGLP----- 1066  
 DB 2134 VSPFKMGFVLNTSAKIECLRGGEWSPSLSVQCIPIVR-----CGEPFELANGYPSGTNYSF 2190  
 QY 1067 --VVVTHSHRKFTDVEVTPGOMT---OYQVLAERAGELGASPPPLNHI--HGAPYCGDGK 1119  
 DB 2191 GAVVAVSCHKGF-----YIKGEKSTCEATQWMSKPTPCHPVSNCNEPKEVNGF 2240  
 QY 1120 VSERLGE-----ECDDGDIVSGD-----GSK----- 1141  
 DB 2241 LEHTTGTFSSEARFQCNPGYKKAAGSVFVCOANRHHNSDAPLSTCLYNGKPPPIONGF 2300  
 QY 1142 -----VCELEBGFNCVGEPSLCYMYEGDICEPERKTSIYDCGIYTPKGY 1187  
 DB 2301 LKGSFEVGSVQGVVC--NEGVELVDGNSTWTCQSGMSKKP--SPKCVPTKCAEPPL 2355  
 QY 1188 LDQWATAYSSHEDKKCPVSLVT--GEPSLICTSYHEDLPNHRPLTGMP-----C 1238  
 DB 2356 ENQVLVLELASE-----VGWMTISCKEGHALQGPVLKCLPBGQ--WNGSFPICKNVLC 2407  
 QY 1239 VASENETQDDRSBQPSGLKKEDVWLKVCNRPGEARAFI-----FLTTGLVPGEHQ 1293  
 DB 2408 -----PSFPL-----IPGVPAAGALHPSGTVKYLVDGFF--LRG 2442  
 QY 1294 QPTVLYLTDVRGNSHSLAGTYGLSCQHNPLIIN-VTHQNVLFHHTTVSLNLFSSPRVGI 1352  
 DB 2443 SPTI-LQADSTWSSPLPECVPECPPEELINGIINVQGLAYLSTLTLYCKPGEFLVG- 2500  
 QY 1353 SAVAL--RTSSRIG--LSAPNSCISDEGONHO-----GO----- 1383  
 DB 2501 NATLCEGNGQWLGKMKPCRIECPREKELINGQFSSVSFOYQGTITYFCDRGRLEGPX 2560  
 QY 1384 --SCIHRPCKQKSCPELLDHDVNVCTSIGP--GLMCA-----ITQORFAL 1429  
 DB 2561 SLTCLB--TGDWMDP--PSCDAHCSDPQPLENFVAGADRYGAMITVSCPFQV 2614  
 QY 1430 QASSGQYIRPMQKEIILLTCSSGHW--DONVCLPYDCGVPD----- 1468  
 DB 2615 LGHAMQ-----TCESGMSSSPTCVPIDGLPHRIDGCTKYRQDGHFQDE 2663  
 QY 1469 -----PSLVYANFSCSGEGTKFL--KCSISCVPAKLOGLSFWLT 1507  
 DB 2664 DMMEEVBYLAHPQHLIETATALENTKESPAASHSHFLYGTWVSYSCEPGEYELGI-FVLI 2722  
 QY 1508 CLEGLWSLEPVVC--KLECDAPRIILANILLPHICQDNNDVGTICKEYCKPGVYVESA 1566  
 DB 2723 CQEDGTWNGTAPSCSTIECDLPVAPENGFL--HFTQTT--MGSAAQYSCKPGHILEGSH 2777  
 QY 1567 EGKVRNKLKIQCLEGGIWEQGS--CIPLYCEPPVPEFG-----MYECTN 1610  
 DB 2778 -----LRL-CLQNKQW--SGTYPRCEALISCSKPNPLNMGSIKGDYSLVLYBSCDS 2827  
 QY 1611 GFSIDSCVLINCQDEREKLPIILTKBGLMTQEFKLCENLOGECPPPSELN----- 1661  
 DB 2828 GYIILNGSKKRTCCGNRD-----WDGHEBMC--IVDCCSPVPVPTNGRYKGEY 2874  
 QY 1662 ---SVEKCKRGY-----GIGAVCSPL-CVLPSPDPVWLPENITADTL 1700  
 DB 2875 FQKEITYSCREGLFLEBAGARSICLTNGSWGAPFSCMPVRCAPVQV-----NGVADGL 2929  
 QY 1701 E-----HMMEPVAVQS--IVCTGRQWHPDPVLAHCIOSECP 1735  
 DB 2930 DYGFKEVAHFCLBEGYVLOGAPRLTCSNGTWDAE-----VVPCKP 2970

RESULT 13  
 US-10-295-027-430  
 ; Sequence 430, Application US/10295027  
 ; Publication No. US2003023350A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Afar, Daniel  
 ; APPLICANT: Aziz, Natasha  
 ; APPLICANT: Ginsberg, Wendy M.

APPLICANT: Gish, Kurt C.  
 APPLICANT: Glynn, Richard  
 APPLICANT: Hevizi, Peter A.  
 APPLICANT: Mack, David H.  
 APPLICANT: Murray, Richard  
 APPLICANT: Watson, Susan R.  
 APPLICANT: Eos Biotechnology, Inc.  
 TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and  
 FILE REFERENCE: 018501-012500US  
 CURRENT APPLICATION NUMBER: US/10/295,027  
 CURRENT FILING DATE: 2002-11-13  
 PRIOR APPLICATION NUMBER: US 09/663,733  
 PRIOR FILING DATE: 2000-09-15  
 PRIOR APPLICATION NUMBER: US 60/350,666  
 PRIOR FILING DATE: 2001-11-13  
 PRIOR APPLICATION NUMBER: US 60/335,394  
 PRIOR FILING DATE: 2001-11-15  
 PRIOR APPLICATION NUMBER: US 60/332,464  
 PRIOR FILING DATE: 2001-11-21  
 PRIOR APPLICATION NUMBER: US 60/334,393  
 PRIOR FILING DATE: 2001-11-29  
 PRIOR APPLICATION NUMBER: US 60/340,376  
 PRIOR FILING DATE: 2001-12-14  
 PRIOR APPLICATION NUMBER: US 60/347,211  
 PRIOR FILING DATE: 2002-01-08  
 PRIOR APPLICATION NUMBER: US 60/347,349  
 PRIOR FILING DATE: 2002-01-10  
 PRIOR APPLICATION NUMBER: US 60/355,250  
 PRIOR FILING DATE: 2002-02-08  
 PRIOR APPLICATION NUMBER: US 60/356,714  
 PRIOR FILING DATE: 2002-02-13  
 Remaining Prior Application data removed - See File Wrapper or PALM.  
 NUMBER OF SEQ ID NOS: 1386  
 SOFTWARE: Patentin Ver. 2.1  
 SEQ ID NO 430  
 LENGTH: 3557  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-10-295-027-430

Query Match 3.4%; Score 334.5; DB 12; Length 3557;  
 Best Local Similarity 19.9%; Pred. No. 1.4e-16;  
 Matches 354; Conservative 207; Mismatches 611; Indels 603; Gaps 106;  
 286 AFTYEAHVKEGCGNNPAIIAGVNDNCST--VSD-KGALGIRSGKDKRARRFFS 341  
 1442 ALTCTFMKSSDDNNYGTPIISYAVDNGSDNTLLTLDYNGWLVY-NGREK----- 1490  
 342 LCTDRVKKATILISHSRVQPGTHTVAATYDGRH--MALVYDGTQVASSLDQSGPLNSPF 399  
 1491 -----ITNCPSVNDRMHIAITMTSANGIMKVYIDGKLSGAGLSVGLPIF- 1538  
 400 MASCRSLILGDSSEDEGHYFR-----GHGLTVFWSTAL-POSHFQSSSHSGSEBEAT 452  
 1539 --GGGALVILGQEDQKKGEGSPASFGSISQLMLMIVLSPQ--QVKSLATSCPEELS 1593  
 453 DLVITASFEPTWTWVPR-----DEK-----YPRL-----EVLQ----- 482  
 1594 KGNVLA-----WPDFLSGIVGKVKIDSKISFCSDCPRLGSGVPHLRTASEDLKGSK 1645  
 483 -----GF-----EPEP-----ELSPLOPL-----CGQTV- 503  
 1646 VNFECDFQVGVNPNVQYCLNQGMQTPPLPFCERISCGVPPPLNGHSHADDFYAGSTVT 1705  
 504 --CDNVELISQYNGYWPJLGEKVIRYOVVNICDEGLNPIVSEBQIRLOHEALNEARSRY 561  
 1706 YQGN-----NGYVILGDSRM-----FCTDNGSMNGVSPSCLOVDECAVSDGSEH 1750  
 562 NISQQLSVHQVHNSITLRRVVLVNCPSKIGN-DHC-DP-ECF--HPLTYDGGDRLQ 615  
 1751 -----ASCLNVDS-----YICSCVPYTGDKNCKAEPICKAKAPGPNENGSHGEIYTV 1799

616 G-----RCYSWNRDGLCHVEC-----NNMLNDFDGDCCDQVADVKTCDPDSPKR 664  
 1800 GAVVTFSCQEGVGLMGTAKTCTGSEBMHLL-----PYCAV--SCCKPAIPEN 1847  
 665 AYSVKELEALONSTHFLNITYFASSVREDLGAATWPDKCAVTHLGIYVSPAYGM 724  
 1848 G--CIEELAFTFGSKTYRCKNYCTIAGDSESCLANSSMSHSP-----VEPVKCS 1899  
 725 PGHTDMIHVGHVGLYHFKGVSRBSCNDCKETVPSEMERGDCIADTA-----P 776  
 1900 PENIN-----GKI-ILSGITLYLSTASVSC-DIGISLQCPSTIECTASGIWDRAP 1948  
 777 TPXSELCREPEPTSDTC--GFTREPGAFTYMSYTDNCTDNFTPNQVAMHCVLDL 832  
 1949 AGLVFCGEPPAKDAVITGNNT-----FRNTVYTT--CEGYTLAQLDITIECLAD- 1998  
 833 VYQQWTEBSK--PTPIPIPMVIGOTNKSLLTHMLPPLSGVVYDRAAGLCACTBEDT 889  
 1999 --GKWSRSDQCLAVSCDEPPIVDHASPE--TAH-----RLFCDIAFYVCSDG- 2042  
 890 FROYVHTASRRVCDSSGVTPEAVGPPD-VQOPCE--PSLQAMSPVHLVHMNMVPC 946  
 2043 --YSLADNSQLCNQKGVVPEGDMPRCIAHFCKEPSVS-----YSI----- 2085  
 947 PTEGCSLELLFQHPVQADTLTLVWTSFFMESQVLPDTIELLE-----NKESVHLGPL 999  
 2086 -----LESYSKAKFAAGS-----VVSFKCMGEVLT-NTSAKICIMCGGNNPBMISQCI 2134  
 1000 DTFCDIPLTIKLVHDGKSVGVYTPDERI-----EIDAALLTSQPHSPICS 2046  
 2135 PVRCGPPPTI--MNGVASGSN--YSFGAMVAVSCNKGFIYKGEKSTCEATGQWMSPTPT 2190  
 1047 GCPRYVQVLRDPPASGLPVVYVTHSHRKTDEVTGQVQVQVLAELAGELGEASPL 2106  
 2191 -CHPV-----SCGEPPKVENGR-----LEHTGRIESEVRYQCPGKYSVSPV 2234  
 1107 -----NHING-APY-----CG-----DGKYSRLGSECDGDLVSGDCS 2293  
 2235 FVCOANRHMHSBPLKCVLPDCKPPIQNGFMKGENFVSGKVOFCNNGYELVGD5-S 2293  
 1141 KVCELEBGFNCVCEPSLCYMYEGDICEPPE-----RKTSYVDGITYT--PK 1185  
 2294 WTCQKSGKNKKKNPK-----CMPAKCEPPLLENQVLKELTTEVGVVTFSCKE 2343  
 1186 GYL-----DOMATRAYSHEDKKKCVSLVTEGPHSLICTSYHDPN--HR 1230  
 2344 GHVLAGPSVLKCLPSQOM-----NDSFPVCKIVLCTPPP--LISFGVPIPSGALHP 2392  
 1231 PLTGMFPCVAS-----ENETODRSEBQEGSLKKEDEVW--LKVCFNRPGEARAFIFLT 1283  
 2393 GSTVTKISCVGPFRLRGNST-----TLQPPDGTMSPLPEC----- 2427  
 1284 TDGLVPEHQOP-TVTLVYLTVDVRSNHSLSGTYSLSQOHN-PLIINYT--HHQNVLFHH 1337  
 2428 -----VPVECPQPEIRPNIIDVQGLAY-LSTALYTCKPGEVLGNTTLTLCGNGHMLGK 2482  
 1338 TTSVLANFSPRKGISAVNALKTSRIGLSAPNSCISEDBEQHOGOS--CHRPCKOD 1394  
 2483 PTCKAIECLKPKKIELNGKESYTDLHYGQVITYSC--NRGFRLEGSSALTCLE--TGDDW 2537  
 1395 -SCPSLLDLHADVAVNCTSIGP--GLMKCA-----ITCGRGFALQASSGQYIRPMQ 1441  
 2538 VDAPS-----CNAIHCDSPPIENGVEGADYSGAIIITYSCPPGVOVAGHAMQ----- 2586  
 1442 KEILLTSSGHWQDQV--SCLPVDGVP-----DPSLVNY----- 1474  
 2587 -----TCESGWSGESSIPTCMPIIDCGLLPPIHIDFBDCTKLKDQGVFEQEDMMIEVPYVTPH 2641  
 1475 -----ANFSCSGCTK-----FL--KRGISCVPPAKXLOGLSBMLTCLDEGLMSLP 1517  
 2642 PPHLAGVAVAKTWENTKESPAHSSNPLVGTWVSYTCNPGVELLG-HPVLLICQEDGTWNGS 2700  
 1518 EYVC-KLECDAPPIIANALLPHCLQDNHDVGTICRYECKPGYVAVASAGKVRNKLK 1576

```

Db 2701 APSCSIECDLPTAPENGFLETFET-----SMGSAVQSCKRGHILAGSD-----LR 2747
Qy 1577 IQCLEGGIWEQGS--CIVNCEPPPPVPEG-----MYECINGSLBQCTLN 1621
Db 2748 L-CLENRKSQASPRCEAISSCKRPVNGSISKSNYYTLSTLYECDPGY-----VLN 2800
Qy 1622 CNOERKXPLILCTEGKLTQEFKLCENTQCECPPESELN-----SVEYKCEQ 1669
Db 2801 GTERR-----TCQDDKXMDDEBPIC--IPVDCSSPVPANQGVNDEYTFQKEIHYTCNE 2853
Qy 1670 GY-----GIGAVCSPL-CVIPSDDPVMLEPENTITADTLEHMEBVK--- 1708
Db 2854 GFLEGAARSRYCLANGSWSGATPDCVPVRCATPP-----QLANGVTBGLDYGFMKEVTPHC 2909
Qy 1709 -----VQSI VTCGRQWHPDPVLVNCIOSCEP 1735
Db 2910 HEGYILHGAPKLTQCSGDNMDAE-----IPLCKP 2938

```

## RESULT 14

```

US-10-295-027-1297
Sequence 1297, Application US/10295027
Publication No. US20030232350A1
GENERAL INFORMATION:
APPLICANT: Afari, Daniel
APPLICANT: Ginsberg, Wendy M.
APPLICANT: Gish, Kurt C.
APPLICANT: Glynn, Richard
APPLICANT: Hevezl, Peter A.
APPLICANT: Mack, David H.
APPLICANT: Murray, Susan R.
APPLICANT: Watson, Richard.
APPLICANT: Eos Biotechnology, Inc.
TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
FILE REFERENCE: 018501-012500US
CURRENT APPLICATION NUMBER: US/10/295,027
CURRENT FILING DATE: 2002-11-13
PRIOR APPLICATION NUMBER: US 09/663,733
PRIOR FILING DATE: 2000-09-15
PRIOR APPLICATION NUMBER: US 60/350,666
PRIOR FILING DATE: 2001-11-13
PRIOR APPLICATION NUMBER: US 60/335,394
PRIOR FILING DATE: 2001-11-15
PRIOR APPLICATION NUMBER: US 60/332,464
PRIOR FILING DATE: 2001-11-21
PRIOR APPLICATION NUMBER: US 60/334,393
PRIOR FILING DATE: 2001-11-29
PRIOR APPLICATION NUMBER: US 60/340,376
PRIOR FILING DATE: 2001-12-14
PRIOR APPLICATION NUMBER: US 60/347,211
PRIOR FILING DATE: 2002-01-08
PRIOR APPLICATION NUMBER: US 60/347,349
PRIOR FILING DATE: 2002-01-10
PRIOR APPLICATION NUMBER: US 60/355,250
PRIOR FILING DATE: 2002-02-08
PRIOR APPLICATION NUMBER: US 60/356,714
PRIOR FILING DATE: 2002-02-13
Remaining Prior Application data removed - See file Wrapper or PALM.
NUMBER OF SEQ ID NOS: 1386
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1297
LENGTH: 3557
TYPE: PRT
ORGANISM: Homo sapiens
US-10-295-027-1297

```

Query Match 3.4%; Score 334.5; DB 12; Length 3557;  
 Best Local Similarity 19.9%; Pred. No. 1.4e-16;  
 Matches 354; Conservative 207; Mismatches 611; Indels 603; Gaps 106;

```

Qy 286 AFTEAVNKEBGGQNNPAILIAGVFNCSHT---VSD-KGMALGIRSGKDKRDAFFFS 341
Db 1442 ALITCFPMKSSDDMNVTPISTAVDNGSDNTLLITDNGVVLVY-NRERK----- 1490
Qy 342 LCTDRVKKATILLISHRYQGTWTHVAATYDGRH--MALVYDGTQVASSLDGSGPLNSPF 399
Db 1491 -----ITNCPENVDRGMHILAITWTSANGKVIWIDGLSDGAGLVGLPIF- 1538
Qy 400 MASGRSLLDGDSSEDDHYFR-----GHGTLVPMSTAL-POSHFOHSSQHSGESEAT 452
Db 1539 --GGGALVLQOBODKKEGFSFPAESFVGSISQNLMDVLSPO---OVKSLATSCPEELS 1593
Qy 453 DLVLTAASFEBVNTWVPR-----DEK-----YPRL-----EVLQ----- 482
Db 1594 KGNVLA-----WRFPLSGIVGKVIDKSIIFCSDCPRLGSGVPHLRTASEDLKRGSK 1645
Qy 483 -----GF-----EPRP-----EILSPLOPL-----CGQTV- 503
Db 1646 VNLFCDBGFQLVGNPVVOYCLNQGQWTOPLPHCERISGCVPPPLENGFHSADDFYAGSTVT 1705
Qy 504 --CDNVELLISQYNGWYPLRGEKVIRYQVNNICDDEGLNPIVSEQRILQHEALNEARSRY 561
Db 1706 YQCN-----NGYLLDGRSM-----FCTDNGSWNGVSP9CLDVBDECAVSGDSEH 1750
Qy 562 NISWQLSVHQVHNSTLRHRVVLVNCBPSKIGN-DHC-DP-ECE--HPLTGYDGCDCRLQ 615
Db 1751 -----ASCNAVDS-----YICSCVPPTYGDGNCAEPIKCAKAPGNBNGHSGEILYTV 1799
Qy 616 G-----RCYSMNRDGLCHVEC-----NNMLNDFDDGDCDPOVADVRKTCFDPDSPKR 664
Db 1800 GAGVTFCGCGYQLMGTKITKICLSESGEMHILI-----PYCKAV--SCGKPAIPEN 1847
Qy 665 AYMSVKELKEALQNSHFNFINIYFASSVREDLAAGATWPMDOXAVHILGILVLSPAYGM 724
Db 1848 G-CIEBLAFTPGSKVYRCKNGKTTLAGDEXSSCLANSSSHSP-----VEBVKCSS 1899
Qy 725 PGHTDTIHEVGHVGLHYHFKVSERESCNDEKETVPMETGDLCADTA-----P 776
Db 1900 PENINN-----GKY-ILSGITLYSTASYS-DIGYSLQGPSIIECTASGIMDRAAP 1948
Qy 777 TPXSELCREBEPISDTC-----GTRFPGAPFTYMSYTDNCTDNFTPNVARNHCYLDL 832
Db 1949 ACHLVFCEBPALIKDAIVTGNFT-----FRMTVYLT--CKEGYTLADLTIECLAD- 1998
Qy 833 VYQWTESRK---PTPIRPMVYIGQTKSLTTHMPLRISGVVYDRASSGLCAGCTEDGT 889
Db 1999 --GKMSRSDDQCLAVSCDEPPIVDHASP--TAH-----RLFGIATFYCSGD- 2042
Qy 890 FRQYVHTASSRRVCDSSGYWTPBEAVGPPD-VDQPCB--PSLOAMSPEVHLYHNMNTVPC 946
Db 2043 ---YSLADNSQLCNAGKKVPRPEGQDMPCIAHFCCKPPSVS-----YSI----- 2085
Qy 947 PTBGSLELLFQHPVQADTTLMTVTSFPMESQVLPFTLEILL-----NKESVHLGPL 999
Db 2086 -----LESVKAFAAGS-----VVSFKMEGFVL-NTSAKIBEMRGQNNPSPMSIQCI 2134
Qy 1000 DTFCDILITLKLHVDKGVSGVKTTPDERI-----EIDALLTSOPHSPLCS 1046
Db 2135 PVRCGEPPSI--NWGASGSN--YSFGAMVAYSCNKFYIKGEKSTICBATQGSSTIPT 2190
Qy 1047 GCRFVRQVLRDPFASGLPVVVTSHRKRFTDVEVTPGMYOYOVLAAGSELEASPL 1106
Db 2191 -CHPV-----SGGEPKVENGR-----LEHTTGRIFESBVRQCNPGYKSVGSFV 2234
Qy 1107 -----NHIG-ADY-----CG-----DGKYSERLGEBCDDGLVSGDGCS 1140
Db 2235 FVCOANRHHMSESLMCPVLDCKGPPRIQNGFMKGEMFEVGSKVQFCNGEYELVGD5-S 2293
Qy 1141 KVCLEBEGFNCVGPSPSLCMTYBEGDIGEPB-----RKTSIVDCGIYT---PK 1185
Db 2294 WTQCKSGKMWKKSNNPK-----CMPAKCEBPRLLENQVLVLELTJTTVGAVTTFSCKE 2343
Qy 1186 GYL-----DQMATRAVSHEDKKKCPVSLVATGPBHLICTSYNPLPN---HR 1230

```

```
Db 2344 GHVLAGSVLKLPSQOW-----NDSFPVCKIVLCTPPP-----LISFGVPIPSALHF 2392
Qy 1231 PLTMPFPCVAS-----ENETODDRSGQREGSLKKEDEW---LKVCPRNRGABARAFIFLT 1283
Db 2293 GSTYKSGCVGFRLRGST-----TLCQPDGWTSSPLPEC-----2427
Qy 1284 TDGLVPEHQP-TVTLVLTVDVRSNHSIAGTYGSLCOHN-PLIINV-----HONVLFPH 1337
Db 2428 ----VPECPQPEELPNCIIVDGLAY-LSALYTCRKGFLVGNNTTLTGSENHMLGSK 2482
Qy 1338 TTSVLNFSRPRVISAVALFTSSRIGLSAPSNCSIEDEGONHOGS---CIHRPCGKOD 1394
Db 2483 PTCRAIBCKEKEIKNGKFSYTDLHYGQTVVYSC---NRGRLGSPSLTCLF--TGDM 2537
Qy 1395 -SCPSLLDHADVNCISGP---GLMKCA-----TICQGFALQASSGQYIRPMQ 1441
Db 2538 VDAIS-----CNAIHCDSPOPTEENGFEAGADYSYGAIITYSQFQVAGHAMQ----- 2586
Qy 1442 KEILLTSSSGHMDQNV-SCLPVDGVP-----DPSLVNY-----1474
Db 2587 ----TCEESGMSSSIPICWIDGCLPRHIDFGDCTKLDQGYFEQEDDMVEVPYVTPH 2641
Qy 1475 ----ANFSCSEGTK-----FL--KRCISICVPAKLGSLSPMLTCLBDGLWSLP 1517
Db 2642 PPHLAGAVAKTMENTKESPATHSSNPLXGTWVSYTCNPGYELLG-NPVLICQEDGTWNGS 2700
Qy 1518 EYVC-KLECDAPIILNANLLPHCLDNHVGITCKECPGYVAESAAGKANKLK 1576
Db 2701 APSGISIECDLPAPENGFLEFET-----SMGSAVOYSCKRGHILAGSD-----LR 2747
Qy 1577 IQCLEGGIMEQGS--CIPVCEPPPPYBEG-----NYECTNGFSLDQCYLN 1621
Db 2748 L-CLBNKMSASRCEAISCKKNPWNAGSIKSNITLYSTLYECDPGI-----VLN 2800
Qy 1622 CNGREKLPILCTYEGMTQSEFKLCEMLQGECPPESELN-----SVEYKCEQ 1669
Db 2801 GTEHR-----TCQDDKNWDEDEPIC--IPVDCSSPPVANGQVGDDEYTFQKEIETVCNE 2853
Qy 1670 GY-----GIGAVCSPL-CYIPSPDPMLEPENTADTLEHMPYK---1708
Db 2854 GFLLEGARSRYCLANGWSGATPDCVPKCATPP---QLANGTEGLDYFMKEVTFHC 2909
Qy 1709 ----VQSVICTGRQWHPDPLVHICIOSCEP 1735
Db 2910 HEGYILHGAAPKLTQOSDGNMDE-----IPLCKP 2938

RESULT 15
US-10-028-248A-8
; Sequence 8, Application US/10028248A
; Publication No. US20030235882A1
; GENERAL INFORMATION:
; APPLICANT: Shimkets, Richard
; APPLICANT: Paturajan, Meera
; APPLICANT: Verner, Corine
; APPLICANT: Casman, Stacie
; APPLICANT: Malyanar, Uriel
; APPLICANT: Shenoy, Suresh
; APPLICANT: Spytek, Kimberly
; APPLICANT: Gangoli, Esna
; APPLICANT: Miller, Charles
; APPLICANT: Boldog, Ferenc
; APPLICANT: Li, Li
; APPLICANT: Taupier Jr, Raymond J
; APPLICANT: Kekuda, Rameah
; APPLICANT: Smithson, Glenda
; APPLICANT: Zerhusen, Bryan
; APPLICANT: Liu, Xiaohong
; APPLICANT: Colman, Steven
; APPLICANT: Tchervnev, Vellizar
; APPLICANT: Si, Jingsheng
; APPLICANT: Edinger, Shlomik
```

```
APPLICANT: Stone, David
APPLICANT: Sciore, Paul
APPLICANT: Miller, Isabelle
APPLICANT: Rothenberg, Mark
TITLE OF INVENTION: No. US20030235882A1el Nucleic Acids and Polypeptides and Methods
FILE REFERENCE: 21402-222
CURRENT APPLICATION NUMBER: US/10/028, 248A
CURRENT FILING DATE: 2001-12-19
PRIORITY APPLICATION NUMBER: 60/256619
PRIORITY FILING DATE: 2000-12-19
PRIORITY APPLICATION NUMBER: 60/262959
PRIORITY FILING DATE: 2001-01-19
PRIORITY APPLICATION NUMBER: 60/272408
PRIORITY FILING DATE: 2001-02-28
PRIORITY APPLICATION NUMBER: 60/285189
PRIORITY FILING DATE: 2001-04-20
PRIORITY APPLICATION NUMBER: 60/308039
PRIORITY FILING DATE: 2001-07-26
PRIORITY APPLICATION NUMBER: 60/311266
NUMBER OF SEQ ID NOS: 211
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 8
LENGTH: 3568
TYPE: PRT
ORGANISM: Homo sapiens
US-10-028-248A-8

Query Match 3.4%; Score 332.5; DB 12; Length 3568;
Best Local Similarity 19.7%; Pred. No. 2e-16;
Matches 352; Conservative 210; Mismatches 602; Indels 619; Gaps 107;

Qy 286 AFTYEANVKEGGONNPAITAGVPDNCSTH---VSD-KGVALGIRSKDYGKRDARFFFS 341
Db 1447 ALTCTFMKMSDDMNVTPTISYAVDNGSDMTLLTLDYNGVLYV-NGREK-----1495
Qy 342 LCTDRVAKATILISHRSYQGTWTHVATY--DGRHMAIYVD-----GTQVASSLDQ 391
Db 1496 -----TNCVSDVNDGRMHILATWTSYGAMRVYINGELSGGTGLSIGKAIIPG 1544
Qy 392 SGPLNSPFMASCSLLIGDSSBDGHYF-----GHIGTLVFWSTAL-PQSHFOSSOH 444
Db 1545 GG-----ALVIGQGDQDKKGEFNPASRVGSIQGLNMDVYLSQ---QVNSLA 1590
Qy 445 SSGEEBATDVLVTRSEPPVTEWPPR-----DEK-----YRL-----478
Db 1591 TSCPEELSKGNVLA-----WPDFLSGIYGVKVIDSKSIFCSDCPLGSGVPHLRTAS 1642
Qy 479 EVLQ-----GP-----EPP-----ELSLPQLPPL-----498
Db 1643 EDLKPSSKVNLFCEPFGQVLGNPNVOYCLNQGQWTPPLPHERCIRCGVPPLENGFHSADD 1702
Qy 499 --CGQTV---CDNVELISQYNGVWPLRGEKVIROYVNIICDDEGLNIVSEBOIRLQHEA 553
Db 1703 FYAGSTVTYQCN-----NGYTLGDSRM-----FCTDNGSMNGVSPSCLDVBECA 1747
Qy 554 LNEAFSRYNISWOLSVHQVNSTLRHRVVLVNCPSKIGN-DHC-DE-ECE---HPLTGY 607
Db 1748 VGSDDCSBH-----ASCLNVDS-----YICSCVPPTYGQKNCABIKKCAPNPENGH 1796
Qy 608 DGDGCRLOQ-----RCYSNRBDGLCHVEC-----NNMNLDPDGGCCPOVADYKTC 656
Db 1797 SSGEITYVGAEVTFSCQEGYOLMGVYKKTITLESGEWMHL-----PYCKAV--SC 1844
Qy 657 FDPSPKRAVSVYELKEALQUNSTHPLNITYFASVYEDLAGAATWPDMDAVTHLGIIV 716
Db 1845 GKPAIPENG--CIBELAFITGSKYTRCNNGYTLLAGKSSCLANSKMSHP-----V 1896
Qy 717 LSPAYYGMGHDTMHEVGHVGLVGVHVFQVSEBESCNDPCKETVPSMETGDLCADTA- 775
Db 1897 CEPVKCSSPENINN-----GKY-ILSGLTYLSTASVSC-DTGVSLQGPSIIECTAS 1945
```

QY 776 -----PTPKSELCEBREPTSDTC---GTFREPGAFPTNYMSYTDNCTDNFTPNQVA 824  
Db 1946 GIMDAPRACHLVPCGEPPAIKDAVITGNMT-----FRNTVYTT---CEGEGTLAGLD 1996  
QY 825 RMHCYLLLVQOQWESRK---PTPIPIPMVIGOTNKSLLTHMLPPLISGVVYDASGSLC 881  
Db 1997 TIECLAD---GKMSRSQOQCLAVSCDEBPPIVDHASP---TAH-----BLFGIDA 2040  
QY 882 GACTEDGTFPROYVHTASRRVCDSSGVWTPPEAAGPPR-VQOQCE---PSLQAMSEVHALY 938  
Db 2041 FYTCSDC---YSLANDSOLICNAQKAVPPEGQDMPCIAHCEKPSVS-----Y 2088  
QY 939 HMMNTVPCPEGSCLELLFOHPVQADTLTLWTSPFMSQVLPDTEILLE-----NK 991  
Db 2089 SI-----LESYSKAKFAAGS---VVSFKMEGFVL-NTSAKICEMRGQWNP 2131  
QY 992 ESVHLGRLDTCDIPLTKLHVQKVGKGVKYYTFDERI-----EIDAAALTS 1038  
Db 2132 SPMSIOCIPIVACGEPPI---MNGYASGSN-YSPGAMVAVSCKNGFYIKGEKSTCEATG 2187  
QY 1039 QPHSPILCSGCRPVRYOVLDRPPASGLPVVYTHSHRKFTDVEVTPGOMYOYQVLAAGGE 1098  
Db 2188 QMSSPDPT-CHPV-----SCGEPKVENGF---LEHTTGRIESEVRYQCNPG 2231  
QY 1099 LGEASPPV-----NHING-APY-----CG-----DGKYSERLGEBCDDGD 1132  
Db 2232 YKSVGSPVFCQANRHMHSSEPLMCVPLDCKPPIQNGFMKGFVGVKQVFCNKG 2291  
QY 1133 LVSGDGSKVELBEGFNCVCEPISLCTVYEGDGICEPFE-----RKTISVDCG 1180  
Db 2292 ELVGBDS-SWTCQKSGKMNKSNPK-----CMPAKCEPPLLENQVLKELYTEVG 2340  
QY 1181 IYT---PKGYL-----DOMATRAYSHEDKKKCPVSLVYGEPSHSLICTSYHPD 1225  
Db 2341 VVTSCKEGHLVQGPVULKCLPSQOM-----NDSFPVCXIVLCTPPP---LISFGVP 2389  
QY 1226 LPN---HRPLTGWPFCVAS---ENETQDRSEQPEGSLKKEDEWV---LKYCFNRPGA 1275  
Db 2390 IPSALHSGSTVKKSCVGFRLRNST-----TLQCPDGTWSSPLPEC-----2432  
QY 1276 RAIFILFTTDLVGEHOQP-TVTLVYLTVDVGSNHSLSGTGLSCQHN-PLIINVT---H 1329  
Db 2433 -----VPECPQPEIPIWGIIIDVOGLAY-ISTALYTCKRPFELVGNNTTLLGE 2479  
QY 1330 HQNVLFHHTTSVLNFSRVRGISAVALRTSSRIGLSAPNSCISEDQNHOGOS---CI 1386  
Db 2480 NGHWLGGKPKCAIECLKPEILNGKESYTDLHGYQVTVYSC---NRGFRLEGPALTCL 2536  
QY 1387 HRPCGKOD-SCPSLLDHADVNNCTSIGP---GLMKCA-----ITCORGALQASS 1433  
Db 2537 E--TGDMVDVAPS-----CNAIHODSPQIENGFEAGADYGAIIIVSCFPGFVAGHA 2589  
QY 1434 GQYIRPMQKEILLTCSSGHMDQNV-SCLPVDCVP-----DPS 1470  
Db 2590 MQ-----TCESBGWSSSIPTCMPIDCGLPHIIFGDCYTKLKDQGYFEQEDDM 2638  
QY 1471 LVNY-----ANFSCSEGT-----FL--KRCGISCVPAKLOGLSPLTCL 1509  
Db 2639 EVPYVTPHPRYHLGVAKTWENTKESPATHSNFIYGTWVSTYCNPGYELUG-NPVLICQ 2697  
QY 1510 EDGLMSLPEVYC-KLECDAPPIILNANILLPHCLQDNHDVGTICKYECCKPGYVVAESAEG 1568  
Db 2698 EDGTMNGSAPSCISIECDLPTAPENGFLRFTET-----SMGSAVOYSCCKPGHILAGSD-- 2750  
QY 1569 KVRNKLKIQLLEGSIWEQGS--CIPVVCBPPPVFEG-----MYECTNGPS 1613  
Db 2751 -----LFL-CLENKMKMGASPRCEAISCCKENPYMNGSIXKSNTYVLSLYECDPGY- 2802  
QY 1614 LDSQCVLNCQBERELPLICTKEGLMTQEFKLCENLQEGECPPPSSELN-----1661  
Db 2803 -----VLNGTERR-----TCODDKNWDDEBPC--IPVDCSSPPVASANGQVRGDEYTPQK 2850  
QY 1662 SVEYKCEQGY-----GIVAVCSPL-CVIPSDDVMLPENITADTLEHM 1703

Db 2851 EIEYTCNEGFILBEGARSRVCLANGSWGATPDVPRACATP-----OLANGVTEGLDYGF 2906  
QY 1704 MEPUK-----VQSIYCTGRQMHDPVVLVHICQSCP 1735  
Db 2907 MKEVTFHCHEGYIILHGA PKLTCOSDGNWDAE-----IPUCKP 2943

Search completed: January 2, 2004, 16:12:09  
Job time : 60 secs



GenCore version 5.1.6  
Copyright: (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2004, 16:02:37 ; Search time 28 Seconds

(without alignments)  
6151.361 Million cell updates/sec

Title: us-09-983-025a-2

Perfect score: 9856  
Sequence: 1 MMCLKIRISLALAGMALC.....ADCDDECTCRDPKAEHQ 1791

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	3916.5	39.7	1627	2	pregnancy-associat
2	292.5	3.0	2489	2	complement C3b/Cab
3	289	2.9	2014	2	complement recepto
4	282.5	2.9	1574	2	MEGf8 protein - ra
5	249	2.5	830	2	P-selectin precurs
6	247	2.5	612	2	B-selectin precurs
7	244	2.5	1091	1	PL0009
8	238.5	2.4	1234	1	NBM5H
9	230.5	2.3	610	2	complement factor
10	228	2.3	551	2	E-selectin precurs
11	227.5	2.3	768	2	endothelial leukoc
12	225	2.3	1394	2	P-selectin - rat
13	224	2.3	1620	2	transforming growt
14	219	2.2	768	2	hypothetical prote
15	218.5	2.2	597	1	P-selectin precurs
16	218.5	2.2	1025	1	C4BP alpha chain p
17	218	2.2	1712	2	complement C3d/Eps
18	217	2.2	5376	2	masking protein pr
19	215	2.2	381	1	zondhesin - mouse
20	215	2.2	440	2	decay-accelerating
21	213.5	2.2	1231	1	decay-accelerating
22	203	2.1	2471	2	complement factor
23	202	2.0	646	2	cell fate determin
24	200	2.0	558	2	complement factor
25	200	2.0	3133	2	P-selectin precurs
26	198	2.0	2437	2	C4BP protein alpha
27	197.5	2.0	610	1	hemocytin - silkw
28	196.5	2.0	1820	1	transmembrane prot
29	196	2.0	485	2	C4b-binding protei
					latent transformin
					E-selectin - bovin

30	196	2.0	1548	2	S34583	serine proteinase
31	196	2.0	2150	2	T32497	hypothetical prote
32	193.5	2.0	597	1	NBRUC4	C4b-binding protei
33	193.5	2.0	2871	2	A55624	fibrinogen-1 precu
34	193	2.0	668	2	A46013	coagulation factor
35	192.5	2.0	2555	2	A40043	notch protein homo
36	191.5	1.9	340	2	I56234	decay-accelerating
37	189	1.9	669	2	S65551	factor H - bovine
38	189	1.9	1429	2	S06434	homocysteine protei
39	189	1.9	3084	1	MMMSA	laminin alpha-1 ch
40	188.5	1.9	808	2	D35069	complement factor
41	188.5	1.9	2871	2	A55567	fibrinogen-1 - bovi
42	188.5	1.9	3051	2	S42373	hypothetical prote
43	188	1.9	1203	2	A49175	Match B protein -
44	187.5	1.9	1111	2	T26972	hypothetical prote
45	187	1.9	482	2	JCS092	E-selectin - pig

## ALIGNMENTS

## RESULT 1

S65464 pregnancy-associated plasma protein A precursor - human

N:Alternate names: PAP-A

C:Species: Homo sapiens (man)

C>Date: 22-Nov-1996 #sequence revision 22-Nov-1996 #text\_change 05-Nov-1999

C/Accession: S65464; S65463; A54220; I38097

R/Haating, J.; Oxvig, C.; Overgaard, M.T.; Ebbesen, P.; Kristensen, T.; Sottrup-Jensen, submitted to the EMBL Data Library, June 1995

A:Description: Complete cDNA sequence of the preproform of human pregnancy-associated p

A:Reference number: S65464

A/Accession: S65464

A/Molecule type: mRNA

A:Residues: 1-1627 <HAA>

A/Cross-References: EMBL:U28727; NID:g1142969; PIDN:AAC50543.1; PID:g1142970

R/Haating, J.; Oxvig, C.; Overgaard, M.T.; Ebbesen, P.; Kristensen, T.; Sottrup-Jensen, Eur. J. Biochem. 237, 159-163, 1996

A:Title: Complete cDNA sequence of the preproform of human pregnancy-associated plasma

A:Reference number: S65463; MUID:96203921; PMID:8620868

A/Accession: S65463

A/Molecule type: mRNA

A:Residues: 1-102 <HAM>

A/Cross-References: EMBL:U28727

A>Note: The authors translated the codon CGA for residue 101 as Thr

R/Kristensen, T.; Oxvig, C.; Sand, O.; Moller, N.P.H.; Sottrup-Jensen, L. Biochemistry 33, 1592-1598, 1994

A:Title: Amino acid sequence of human pregnancy-associated plasma protein-A derived from

A:Reference number: A54220; MUID:94146014; PMID:7508748

A/Accession: A54220

A/Molecule type: mRNA

A:Residues: 77-1627 <KRI>

A/Cross-References: GB:X68280; NID:g394649; PIDN:CAA48341.1; PID:g394650

R/Oxvig, C.; Sand, O.; Kristensen, T.; Gleich, G.J.; Sottrup-Jensen, L. J. Biol. Chem. 269, 12243-12246, 1993

A:Title: Circulating human pregnancy-associated plasma protein-A is disulfide-bridged to

A:Reference number: I38097; MUID:93286645; PMID:7685339

A/Accession: I38097

A/Status: translated from GB/EMBL/DDBJ

A/Molecule type: mRNA

A:Residues: 77-1627 <RES>

A/Cross-References: EMBL:X68280; NID:g394649; PIDN:CAA48341.1; PID:g394650

C/Genetic: GDB:134729; OMIM:176385

A:Gene: GDB:134729; OMIM:176385

A/Cross-References: GDB:134729; OMIM:176385

A/Map position: 9q33.1-9q33.1

F:1-22/Domain: signal sequence #status predicted <SIG>

F:23-80/Domain: propeptide #status predicted <PRO>

F:81-1627/Product: pregnancy-associated plasma protein A #status predicted <MAT>

Query Match 39.7%; Score 3916.5; DB 2; Length 1627;  
Best Local Similarity 45.8%; Pred. No. 1.8e-246;  
Matches 718; Conservative 296; Mismatches 499; Indels 53; Gaps 25;

**June**

## RESULT 2

## RESULT 2

complement C3b/C4b receptor, membrane-bound form precursor - human  
N.Alternate names: complement C3b/C4b receptor; complement receptor type 1 (CRI); surfactant  
N.Contains: complement C3b/C4b receptor, secreted form  
C.Species: Homo sapiens (man)  
C.Date: 02-Jul-1996 #sequence revision 24-Nov-1999 #ext. change 21-Jul-2000  
C.Accession: I73012; I55203; A47602; S03291; S03843; A28507; A24748; B24748  
R.Viik, D.P.; Wong, W.W.  
J. Immunol. 151, 6214-6224, 1993  
A.Title: Structure of the gene for the F allele of complement receptor type 1 and sequen  
A.Reference number: I56203; MUID:94055175; PMID:8245463  
A.Accession: I73012  
A.Status: preliminary; translated from GB/EMBL/DBD  
A.Molecule type: DNA  
A.Residues: 1-683, 'X', 1685-1133, 'X', 1135-1471, 'X', 1473-2489 <VIK1>  
A.Cross-references: GB:I17418; NID:G306678; PION:AAB60695.1; PID:G451203  
A.Accession: I56203  
A.Status: translated from GB/EMBL/DBD  
A.Molecule type: DNA  
A.Residues: 1-683, 'X', 685-994, 'A', 896-1000, 1451-1471, 'X', 1473-2489 <VIK2>  
A.Cross-references: GB:I17418; NID:G306678; PION:AAB60694.1; PID:G306680  
J.Wong, W.W.; Cahill, J.M.; Rosen, M.D.; Kennedy, C.A.; Bonocito, E.T.; Morris, M.J.; Wi  
J. Exp. Med. 169, 847-863, 1989  
A.Title: Structure of the human CRI gene: Molecular basis of the structural and quantita  
A.Reference number: A47602; MUID:89176869; PMID:2564414  
A.Accession: A47602  
A.Molecule type: DNA  
A.Residues: 1-41 <KON>  
A.Cross-references: GB:X14893  
R.Hourcade, D.; Maesner, D.R.; Atkinson, J.P.; Holers, V.M.  
J. Exp. Med. 168, 1255-1270, 1988  
A.Title: Identification of an alternative polyadenylation site in the human C3b/C4b recep  
type 1.  
A.Reference number: S03291; MUID:89010527; PMID:2971757  
A.Accession: S03291

```

QY 61 ASPOHHLFGVYPSRAGNYLRPYVGEQOBIHHTGRSKPTDEGNVAVSLVPEPDLTENPAGLRG 120
DB 61 ASPOHHLFGVYPSRAGNYLRPYVGEQOBIHHTGRSKPTDEGNVAVSLVPEPDLTENPAGLRG 120
QY 121 AVEEPAAMVWGDSPIGOSELLGDDDAIYGNORSKSLGEBAGI QKSSAMAAATTTTIAFTTL 180
DB 121 AVEEPAAMVWGDSPIGOSELLGDDDAIYGNORSKSLGEBAGI QKSSAMAAATTTTIAFTTL 180
QY 121 AVEEPAAMVWGDSPIGOSELLGDDDAIYGNORSKSLGEBAGI QKSSAMAAATTTTIAFTTL 180
DB 121 AVEEPAAMVWGDSPIGOSELLGDDDAIYGNORSKSLGEBAGI QKSSAMAAATTTTIAFTTL 180
QY 121 NEPKETORRGWAKSROKRWKRAEDQGGSGISSHFQPMFKSLKRWKSKSPPEESN 240
DB 121 NEPKETORRGWAKSROKRWKRAEDQGGSGISSHFQPMFKSLKRWKSKSPPEESN 240
QY 121 NEPKETORRGWAKSROKRWKRAEDQGGSGISSHFQPMFKSLKRWKSKSPPEESN 240
DB 121 NEPKETORRGWAKSROKRWKRAEDQGGSGISSHFQPMFKSLKRWKSKSPPEESN 240
QY 241 QNGGSGYREAEFTFNSOVGLPILYFSGRERELLRPEVLAELPREAFTVEAMVKEGQON 300
DB 241 QNGGSGYREAEFTFNSOVGLPILYFSGRERELLRPEVLAELPREAFTVEAMVKEGQON 300
QY 241 QNGGSGYREAEFTFNSOVGLPILYFSGRERELLRPEVLAELPREAFTVEAMVKEGQON 300
DB 241 QNGGSGYREAEFTFNSOVGLPILYFSGRERELLRPEVLAELPREAFTVEAMVKEGQON 300
QY 301 NPATLACVFDNCSHVSVDKGMALGIRSGDKGRADAPFPLSLCTRYKATTLISHSRQ 360
DB 301 NPATLACVFDNCSHVSVDKGMALGIRSGDKGRADAPFPLSLCTRYKATTLISHSRQ 360
QY 301 NPATLACVFDNCSHVSVDKGMALGIRSGDKGRADAPFPLSLCTRYKATTLISHSRQ 360
DB 301 NPATLACVFDNCSHVSVDKGMALGIRSGDKGRADAPFPLSLCTRYKATTLISHSRQ 360
QY 361 PGTWTHVATYDGRHMAIYVDGNOVASSLDOSGPLNSPMAASCSLLGSDSSBDGHPYR 420
DB 361 PGTWTHVATYDGRHMAIYVDGNOVASSLDOSGPLNSPMAASCSLLGSDSSBDGHPYR 420
QY 361 PGTWTHVATYDGRHMAIYVDGNOVASSLDOSGPLNSPMAASCSLLGSDSSBDGHPYR 420
DB 361 PGTWTHVATYDGRHMAIYVDGNOVASSLDOSGPLNSPMAASCSLLGSDSSBDGHPYR 420
QY 421 GHLGLTVWSTALPOSHFQSSQHSGBEEDTLVLTASFEVNTWVPEFDEKYPRLV 480
DB 421 GHLGLTVWSTALPOSHFQSSQHSGBEEDTLVLTASFEVNTWVPEFDEKYPRLV 480
QY 421 GHLGLTVWSTALPOSHFQSSQHSGBEEDTLVLTASFEVNTWVPEFDEKYPRLV 480
DB 421 GHLGLTVWSTALPOSHFQSSQHSGBEEDTLVLTASFEVNTWVPEFDEKYPRLV 480
QY 481 LOGFEPEPEPILPLQPLCGQTYCQNVNELISQYNGYWPRLRGKXVIRYQVNI CDEGLNP 540
DB 481 LOGFEPEPEPILPLQPLCGQTYCQNVNELISQYNGYWPRLRGKXVIRYQVNI CDEGLNP 540
QY 481 LOGFEPEPEPILPLQPLCGQTYCQNVNELISQYNGYWPRLRGKXVIRYQVNI CDEGLNP 540
DB 481 LOGFEPEPEPILPLQPLCGQTYCQNVNELISQYNGYWPRLRGKXVIRYQVNI CDEGLNP 540
QY 541 IYSEEOIRIQHEALNAFSTRYNI SMQSYHOVNSTLRHRYLVNCEPEKIGNDHDCPPC 600
DB 541 IYSEEOIRIQHEALNAFSTRYNI SMQSYHOVNSTLRHRYLVNCEPEKIGNDHDCPPC 600
QY 541 IYSEEOIRIQHEALNAFSTRYNI SMQSYHOVNSTLRHRYLVNCEPEKIGNDHDCPPC 600
DB 541 IYSEEOIRIQHEALNAFSTRYNI SMQSYHOVNSTLRHRYLVNCEPEKIGNDHDCPPC 600
QY 601 EHPPLTGYDGGDRLQGRCTYSWNRDGLCHVECNMNLNDFDDGDCDPOVADVAKTCFDD 660
DB 601 EHPPLTGYDGGDRLQGRCTYSWNRDGLCHVECNMNLNDFDDGDCDPOVADVAKTCFDD 660
QY 601 EHPPLTGYDGGDRLQGRCTYSWNRDGLCHVECNMNLNDFDDGDCDPOVADVAKTCFDD 660
DB 601 EHPPLTGYDGGDRLQGRCTYSWNRDGLCHVECNMNLNDFDDGDCDPOVADVAKTCFDD 660
QY 661 SPKRAVMSYKELKALQNLSTHFLNIPASSVREDLAAGATWMDAATVHJGLVLSA 720
DB 661 SPKRAVMSYKELKALQNLSTHFLNIPASSVREDLAAGATWMDAATVHJGLVLSA 720
QY 661 SPKRAVMSYKELKALQNLSTHFLNIPASSVREDLAAGATWMDAATVHJGLVLSA 720
DB 661 SPKRAVMSYKELKALQNLSTHFLNIPASSVREDLAAGATWMDAATVHJGLVLSA 720
QY 721 YYGMEGHTDMHEVHVLGLYHVFQVSEBESCNBPCKEYVSMETGDLCAJTAFTPKS 780
DB 721 YYGMEGHTDMHEVHVLGLYHVFQVSEBESCNBPCKEYVSMETGDLCAJTAFTPKS 780
QY 721 YYGMEGHTDMHEVHVLGLYHVFQVSEBESCNBPCKEYVSMETGDLCAJTAFTPKS 780
DB 721 YYGMEGHTDMHEVHVLGLYHVFQVSEBESCNBPCKEYVSMETGDLCAJTAFTPKS 780
QY 781 ELCREBEPTSDTCGFTFPGAPFTNTMSTYTDNCTDNFTPNQVARNHCTLDLVYQOMTES 840
DB 781 ELCREBEPTSDTCGFTFPGAPFTNTMSTYTDNCTDNFTPNQVARNHCTLDLVYQOMTES 840
QY 781 ELCREBEPTSDTCGFTFPGAPFTNTMSTYTDNCTDNFTPNQVARNHCTLDLVYQOMTES 840
DB 781 ELCREBEPTSDTCGFTFPGAPFTNTMSTYTDNCTDNFTPNQVARNHCTLDLVYQOMTES 840
QY 841 RKPRTPIRPMYIGONKSLITHTMLPIGIVVYDRAAGSICGACTDGTFRQVHTAASR 900
DB 841 RKPRTPIRPMYIGONKSLITHTMLPIGIVVYDRAAGSICGACTDGTFRQVHTAASR 900
QY 841 RKPRTPIRPMYIGONKSLITHTMLPIGIVVYDRAAGSICGACTDGTFRQVHTAASR 900
DB 841 RKPRTPIRPMYIGONKSLITHTMLPIGIVVYDRAAGSICGACTDGTFRQVHTAASR 900
QY 901 RUCDSGSGYTPREAVPDPVDCPCEBSLOAMSPEVLYHNMNVVPCPTGCSJELLFOHP 960
DB 901 RUCDSGSGYTPREAVPDPVDCPCEBSLOAMSPEVLYHNMNVVPCPTGCSJELLFOHP 960
QY 901 RUCDSGSGYTPREAVPDPVDCPCEBSLOAMSPEVLYHNMNVVPCPTGCSJELLFOHP 960
DB 901 RUCDSGSGYTPREAVPDPVDCPCEBSLOAMSPEVLYHNMNVVPCPTGCSJELLFOHP 960
QY 961 VOADTLTLWVTSFPMESQVLFTEILLNKESVHJGLPDTFCIDPLITLHJADGKVSQV 1020
DB 961 VOADTLTLWVTSFPMESQVLFTEILLNKESVHJGLPDTFCIDPLITLHJADGKVSQV 1020
QY 961 VOADTLTLWVTSFPMESQVLFTEILLNKESVHJGLPDTFCIDPLITLHJADGKVSQV 1020
DB 961 VOADTLTLWVTSFPMESQVLFTEILLNKESVHJGLPDTFCIDPLITLHJADGKVSQV 1020
QY 1021 KYTTPERIEIDALLTSOPHSPLGSGCPVRYOVLADPPAAGLGVVYVTHSHRKTDE 1080
DB 1021 KYTTPERIEIDALLTSOPHSPLGSGCPVRYOVLADPPAAGLGVVYVTHSHRKTDE 1080
QY 1021 KYTTPERIEIDALLTSOPHSPLGSGCPVRYOVLADPPAAGLGVVYVTHSHRKTDE 1080
DB 1021 KYTTPERIEIDALLTSOPHSPLGSGCPVRYOVLADPPAAGLGVVYVTHSHRKTDE 1080
QY 1081 VTPGQMYOYOVLAAGGELGEASPLNHIHGAFCYCGDKVSESLGECDDGDVSDGGS 1140
DB 1081 VTPGQMYOYOVLAAGGELGEASPLNHIHGAFCYCGDKVSESLGECDDGDVSDGGS 1140
QY 1081 VTPGQMYOYOVLAAGGELGEASPLNHIHGAFCYCGDKVSESLGECDDGDVSDGGS 1140
DB 1081 VTPGQMYOYOVLAAGGELGEASPLNHIHGAFCYCGDKVSESLGECDDGDVSDGGS 1140
QY 1141 KVCLEBEGNVCPEPBLCTMYEGDGI CEPPERKTSIYDCGIIYTPKGYLQDMATRAYSSHE 1200
DB 1141 KVCLEBEGNVCPEPBLCTMYEGDGI CEPPERKTSIYDCGIIYTPKGYLQDMATRAYSSHE 1200

```

```

DB 1141 KVCLEBEGNVCPEPBLCTMYEGDGI CEPPERKTSIYDCGIIYTPKGYLQDMATRAYSSHE 1200
QY 1201 DKKCCPVSLVTGEBHSILCTSYHDDLPHNPLTCWPCVASENTOODRSOEPGSLKE 1260
DB 1201 DKKCCPVSLVTGEBHSILCTSYHDDLPHNPLTCWPCVASENTOODRSOEPGSLKE 1260
QY 1201 DKKCCPVSLVTGEBHSILCTSYHDDLPHNPLTCWPCVASENTOODRSOEPGSLKE 1260
DB 1201 DKKCCPVSLVTGEBHSILCTSYHDDLPHNPLTCWPCVASENTOODRSOEPGSLKE 1260
QY 1261 DEVALKVCENRPGBARAIPIFLTTDGLVPGHQPTVTLVLTDRGSHSISGTYGLSCOH 1320
DB 1261 DEVALKVCENRPGBARAIPIFLTTDGLVPGHQPTVTLVLTDRGSHSISGTYGLSCOH 1320
QY 1261 DEVALKVCENRPGBARAIPIFLTTDGLVPGHQPTVTLVLTDRGSHSISGTYGLSCOH 1320
DB 1261 DEVALKVCENRPGBARAIPIFLTTDGLVPGHQPTVTLVLTDRGSHSISGTYGLSCOH 1320
QY 1321 NPILLINTHONVLFHHTSYLNFSSPRGICSAVALRTSRIGLSAPNSCISDDEQON 1380
DB 1321 NPILLINTHONVLFHHTSYLNFSSPRGICSAVALRTSRIGLSAPNSCISDDEQON 1380
QY 1321 NPILLINTHONVLFHHTSYLNFSSPRGICSAVALRTSRIGLSAPNSCISDDEQON 1380
DB 1321 NPILLINTHONVLFHHTSYLNFSSPRGICSAVALRTSRIGLSAPNSCISDDEQON 1380
QY 1381 OGSCCHRPCCGKODSCSLLDHDVYVCTSIGGLMKCAITTCORGFALQASGOYIRPM 1440
DB 1381 OGSCCHRPCCGKODSCSLLDHDVYVCTSIGGLMKCAITTCORGFALQASGOYIRPM 1440
QY 1441 OKEILLITCSGHWQONVSCLPVDCGVPDPSLVNYANFSCSEGTFLKRCISISVPPAKLQ 1500
DB 1441 OKEILLITCSGHWQONVSCLPVDCGVPDPSLVNYANFSCSEGTFLKRCISISVPPAKLQ 1500
QY 1441 OKEILLITCSGHWQONVSCLPVDCGVPDPSLVNYANFSCSEGTFLKRCISISVPPAKLQ 1500
DB 1441 OKEILLITCSGHWQONVSCLPVDCGVPDPSLVNYANFSCSEGTFLKRCISISVPPAKLQ 1500
QY 1501 GLSPMLTCLBEGDGLMSLBEVYCKECDAPITILANLPLHCLQDNHVGITCKYCKRGY 1560
DB 1501 GLSPMLTCLBEGDGLMSLBEVYCKECDAPITILANLPLHCLQDNHVGITCKYCKRGY 1560
QY 1501 GLSPMLTCLBEGDGLMSLBEVYCKECDAPITILANLPLHCLQDNHVGITCKYCKRGY 1560
DB 1501 GLSPMLTCLBEGDGLMSLBEVYCKECDAPITILANLPLHCLQDNHVGITCKYCKRGY 1560
QY 1561 YVARSABEKVANKLKTQCLEGGIMBQSCIPVCEBPPEVPEGMECTGNSFSDSOCLV 1620
DB 1561 YVARSABEKVANKLKTQCLEGGIMBQSCIPVCEBPPEVPEGMECTGNSFSDSOCLV 1620
QY 1561 YVARSABEKVANKLKTQCLEGGIMBQSCIPVCEBPPEVPEGMECTGNSFSDSOCLV 1620
DB 1561 YVARSABEKVANKLKTQCLEGGIMBQSCIPVCEBPPEVPEGMECTGNSFSDSOCLV 1620
QY 1621 NONGERELPILCTKEGLMTQEPFLCENTLOCECPPEPSELNSVYKCOGIGIVACPL 1680
DB 1621 NONGERELPILCTKEGLMTQEPFLCENTLOCECPPEPSELNSVYKCOGIGIVACPL 1680
QY 1621 NONGERELPILCTKEGLMTQEPFLCENTLOCECPPEPSELNSVYKCOGIGIVACPL 1680
DB 1621 NONGERELPILCTKEGLMTQEPFLCENTLOCECPPEPSELNSVYKCOGIGIVACPL 1680
QY 1681 CYPSPSDVMLPENITADTLTHMMEPVYVOSIVCTGRQOMHPDPLVHCISCEPFOADG 1740
DB 1681 CYPSPSDVMLPENITADTLTHMMEPVYVOSIVCTGRQOMHPDPLVHCISCEPFOADG 1740
QY 1681 CYPSPSDVMLPENITADTLTHMMEPVYVOSIVCTGRQOMHPDPLVHCISCEPFOADG 1740
DB 1681 CYPSPSDVMLPENITADTLTHMMEPVYVOSIVCTGRQOMHPDPLVHCISCEPFOADG 1740
QY 1741 WCDTINRAYCHYDGGDCSSSTLSKVIYPPAADCDDDECTCRPKAENQ 1791
DB 1741 WCDTINRAYCHYDGGDCSSSTLSKVIYPPAADCDDDECTCRPKAENQ 1791
QY 1741 WCDTINRAYCHYDGGDCSSSTLSKVIYPPAADCDDDECTCRPKAENQ 1791
DB 1741 WCDTINRAYCHYDGGDCSSSTLSKVIYPPAADCDDDECTCRPKAENQ 1791

```

## RESULT 2

US-09-827-998-3

Sequence 3, Application US/09827998

Patent No. US2002010252A1

GENERAL INFORMATION:

APPLICANT: Gu, Yizhong

APPLICANT: Shannon, Mark

TITLE OF INVENTION: NOVEL ISOFORMS OF HUMAN PREGNANCY-ASSOCIATED PROTEIN B

FILE REFERENCE: MDHMR-8

CURRENT FILING DATE: 2001-04-06

PRIOR APPLICATION NUMBER: US 60/207,456

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: US 60/236,359

PRIOR FILING DATE: 2000-09-27

NUMBER OF SEQ ID NOS: 1881

SOFTWARE: Aecmica Sequence Listing Engine

SEQ ID NO: 3

LENGTH: 1791

TYPE: PRT

ORGANISM: Homo sapiens

US-09-827-998-3

Query Match 99.8%; Score 9836; DB 10; Length 1791;  
 Best Local Similarity 99.8%; Pred. No. 0;  
 Matches 1788; Conservative 2; Indels 0; Gaps 0;  
 1 MNCILIRISLAILAGNALCSANSELGWTTRKSLYERHINOVLLGGRCTGAVRPR 60

Db 1 MMCLILRLISLAILAGMALCSANSELGWTTRKSLVEREHLNQLLEGERCWLGAARBR 60  
 Qy 61 ASPOHHLFGVYPSQAGNLYRPPVGEQEIHTHGRKPTTEGNAVSLVPPDLTENPARG 120  
 Db 61 ASPOHHLFGVYPSQAGNLYRPPVGEQEIHTHGRKPTTEGNAVSLVPPDLTENPARG 120  
 Qy 121 AVEERPAWVDSPLIGSELGLDADAYLGNORSKSLBAGIOKSAMAATTITLFTTL 180  
 Db 121 AVEERPAWVDSPLIGSELGLDADAYLGNORSKSLBAGIOKSAMAATTITLFTTL 180  
 Qy 181 NEPKETORRGWAKSRORROYWKRAEDQDGSISSEHQPWKSLKXRVKSPSEBN 240  
 Db 181 NEPKETORRGWAKSRORROYWKRAEDQDGSISSEHQPWKSLKXRVKSPSEBN 240  
 Qy 241 ONGEGSYREAEFTNSQVGLPLLYSGRERILLPREVLAELPREAFTEAVKPEGON 300  
 Db 241 ONGEGSYREAEFTNSQVGLPLLYSGRERILLPREVLAELPREAFTEAVKPEGON 300  
 Qy 301 NPAILAGVFNDSHVSVDKGMALGIRSGDKGRDRPFPSLCTDRVKKATLISHSRQ 360  
 Db 301 NPAILAGVFNDSHVSVDKGMALGIRSGDKGRDRPFPSLCTDRVKKATLISHSRQ 360  
 Qy 361 PCTWTHVATYDGRHVALYVDGTQVASSIDQSGPLNSPMAACRSILLGDSSEDEHYR 420  
 Db 361 PCTWTHVATYDGRHVALYVDGTQVASSIDQSGPLNSPMAACRSILLGDSSEDEHYR 420  
 Qy 421 GHLGTLVFWSTALPOSHFOHSSGSEBEATDLVLTASFEPTNTEWVPEDEKRYRLV 480  
 Db 421 GHLGTLVFWSTALPOSHFOHSSGSEBEATDLVLTASFEPTNTEWVPEDEKRYRLV 480  
 Qy 481 LOGFEDEPILSPLOPPLCGQTCVNVIELISQYNGWPLGEKVIYQVNNICDEGLNP 540  
 Db 481 LOGFEDEPILSPLOPPLCGQTCVNVIELISQYNGWPLGEKVIYQVNNICDEGLNP 540  
 Qy 541 IYSEBQIRLOHEALNEAFSRYNISWQLSVHQNSTLRHVRVLCNEPSKIGNDHCPBC 600  
 Db 541 IYSEBQIRLOHEALNEAFSRYNISWQLSVHQNSTLRHVRVLCNEPSKIGNDHCPBC 600  
 Qy 601 EHPPLTGYDGGDCLQGRCTISNRDDGICHECNMMLNDPDDGCCPOVADVAKTCEDP 660  
 Db 601 EHPPLTGYDGGDCLQGRCTISNRDDGICHECNMMLNDPDDGCCPOVADVAKTCEDP 660  
 Qy 661 SPBRAYSVKELKALQUNSTHFLNTFYASSVEDLAGAATWPMKDQAVTHLGSIVSPA 720  
 Db 661 SPBRAYSVKELKALQUNSTHFLNTFYASSVEDLAGAATWPMKDQAVTHLGSIVSPA 720  
 Qy 721 YYGMPGHTDTMIEHGVHLGLYHVFKGVSERESCNDCCKETVPSMETGDLCAUTAIPPKS 780  
 Db 721 YYGMPGHTDTMIEHGVHLGLYHVFKGVSERESCNDCCKETVPSMETGDLCAUTAIPPKS 780  
 Qy 781 ELCREPEPISDTGCTFRPGAPFTNYSYTDNCTDNFTNOVARMHCYLDIVYQOMTES 840  
 Db 781 ELCREPEPISDTGCTFRPGAPFTNYSYTDNCTDNFTNOVARMHCYLDIVYQOMTES 840  
 Qy 841 RKCPPIPIPMVNGQTKSLTIHMLPLISGVVYRASGSLCAGCTBGTFRQVYHTASSR 900  
 Db 841 RKCPPIPIPMVNGQTKSLTIHMLPLISGVVYRASGSLCAGCTBGTFRQVYHTASSR 900  
 Qy 901 RVCDSSGYMPEEAAGPPDVDOCEPSLOANSPEVHLVYNNMTVPCPTGCSLELFOHP 960  
 Db 901 RVCDSSGYMPEEAAGPPDVDOCEPSLOANSPEVHLVYNNMTVPCPTGCSLELFOHP 960  
 Qy 961 VOADTLTLMVTSFPMESQVLPTEIILENKESVHLGLDPCDIPILIKLAVDGKSGV 1020  
 Db 961 VOADTLTLMVTSFPMESQVLPTEIILENKESVHLGLDPCDIPILIKLAVDGKSGV 1020  
 Qy 1021 KVTYFDEIEIIDAALLTQSPHSPICSGGRPVRYOVLAPPPASGLPVVVTSHRKFTVVE 1080  
 Db 1021 KVTYFDEIEIIDAALLTQSPHSPICSGGRPVRYOVLAPPPASGLPVVVTSHRKFTVVE 1080  
 Qy 1081 VTPGOMYOYVLAAGELGEASPLNHIHGAFCYGDGKVSERLGECDDDIVSGDGS 1140  
 Db 1081 VTPGOMYOYVLAAGELGEASPLNHIHGAFCYGDGKVSERLGECDDDIVSGDGS 1140

Qy 1141 KVCELEEGFNCVGEPSLCYNTBQDGTCEPPEKRTSYVDGCIYTPKGYLDQMATRAVSSHE 1200  
 Db 1141 KVCELEEGFNCVGEPSLCYNTBQDGTCEPPEKRTSYVDGCIYTPKGYLDQMATRAVSSHE 1200  
 Qy 1201 DKKKCPVSLYGEBSHSLICTSYHDPDLPHNRPLTGMPCVASENETODDRSEOPESLAKKE 1260  
 Db 1201 DKKKCPVSLYGEBSHSLICTSYHDPDLPHNRPLTGMPCVASENETODDRSEOPESLAKKE 1260  
 Qy 1261 DEWMLKVCFNRPGEARAFIFLITDGLVPGHQPTVLYLTVDRGSNHSIGTYGLSCOH 1320  
 Db 1261 DEWMLKVCFNRPGEARAFIFLITDGLVPGHQPTVLYLTVDRGSNHSIGTYGLSCOH 1320  
 Qy 1321 NPLIINTVTHQNVLFHHTTSVLNPSSEPRVIGSAVALRTSSRIGLSAPSNCTISDEGONH 1380  
 Db 1321 NPLIINTVTHQNVLFHHTTSVLNPSSEPRVIGSAVALRTSSRIGLSAPSNCTISDEGONH 1380  
 Qy 1381 OGOSCIHRPCGKQSCPSLLLDHADVNCSTIGELMKCAITTCORGALQASGQYIRPM 1440  
 Db 1381 OGOSCIHRPCGKQSCPSLLLDHADVNCSTIGELMKCAITTCORGALQASGQYIRPM 1440  
 Qy 1441 OKELILTCSSGHWQNVCLPVDGVPDPSLVNANFSCSBGTFLKRCISICVPAKQ 1500  
 Db 1441 OKELILTCSSGHWQNVCLPVDGVPDPSLVNANFSCSBGTFLKRCISICVPAKQ 1500  
 Qy 1501 GLSPWLTCLEBDGLMSLPEVYCKLBCDAPPIIINANLILPHCLQDNDVGTCKYCKRXY 1560  
 Db 1501 GLSPWLTCLEBDGLMSLPEVYCKLBCDAPPIIINANLILPHCLQDNDVGTCKYCKRXY 1560  
 Qy 1561 YVASEABGVNRKLIKICOLEGIMEQSCIPVCEPPPVBEVGEYCTNGFSLDQCVL 1620  
 Db 1561 YVASEABGVNRKLIKICOLEGIMEQSCIPVCEPPPVBEVGEYCTNGFSLDQCVL 1620  
 Qy 1621 NCNOERKPLICTKEGLMTOEFKLCENLOQSCPPPSLSNVRKCEQYIGAVCSPL 1680  
 Db 1621 NCNOERKPLICTKEGLMTOEFKLCENLOQSCPPPSLSNVRKCEQYIGAVCSPL 1680  
 Qy 1681 CVIPSDPVMLENTITADTLTLEHMEBVKYQSVCTGRNRWMDPVLVHCIOGCEPQANG 1740  
 Db 1681 CVIPSDPVMLENTITADTLTLEHMEBVKYQSVCTGRNRWMDPVLVHCIOGCEPQANG 1740  
 Qy 1741 WCDTINNRVCHYDGGDCSSSTLSKCVIPFAADCDLDECTCRDPAEENQ 1791  
 Db 1741 WCDTINNRVCHYDGGDCSSSTLSKCVIPFAADCDLDECTCRDPAEENQ 1791

RESULT 3  
 US-09-827-998-10  
 ; Sequence 10, Application US/09827998  
 ; Patent No. US20020102252n1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gu. Yizhong  
 ; APPLICANT: Shannon, Mark  
 ; TITLE OF INVENTION: NOVEL ISOFORMS OF HUMAN PREGNANCY-ASSOCIATED PROTEIN E  
 ; FILE REFERENCE: MDHMR-8  
 ; CURRENT APPLICATION NUMBER: US/09/827,998  
 ; PRIOR FILING DATE: 2001-04-06  
 ; PRIOR APPLICATION NUMBER: US 60/207,456  
 ; PRIOR FILING DATE: 2000-05-26  
 ; PRIOR APPLICATION NUMBER: US 60/236,359  
 ; PRIOR FILING DATE: 2000-09-27  
 ; NUMBER OF SEQ ID NOS: 1881  
 ; SOFTWARE: Acomica Sequence Listing Engine  
 ; SEQ ID NO 10  
 ; LENGTH: 1770  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; US-09-827-998-10  
 Query Match 96.54; Score 9507; DB 10; Length 1770;  
 Best Local Similarity 99.94; Pred. No. 0;  
 Matches 1732; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

InterPro: IPR000436; Sushi SCR CCP.  
 DR InterPro: IPR006025; Zn\_MTPdepcase.  
 DR Pfam: PF000084; sushi; 4.  
 DR SMART: SM00032; CCP; 4.  
 DR SMART: SM00004; NL; 2.  
 DR PROSITE: PS00092; N6\_MTASB; 1.  
 DR PROSITE: PS00142; ZINC\_PROTEASE; 1.  
 DR SEQUENCE 1790 AA; 158450 MW; E40717B65623A0E9 CRC64;

Query Match 99.7%; Score 9831; DB 4; Length 1790;  
 Best Local Similarity 99.8%; Pred. No. 0;  
 Matches 1787; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

2 MCLTILSLAIIAGMALCSANSELGWTBKSLVERELNOVLEGERCMLGAVRPRRA 61  
 1 MCLTILSLAIIAGMALCSANSELGWTBKSLVERELNOVLEGERCMLGAVRPRRA 60  
 62 SPOHHLFGVYPSRAGNYLRPYVGEQEIHTGRSKPDTGNAVSLVPPDLTENPAGLGA 121  
 61 SPOHHLFGVYPSRAGNYLRPYVGEQEIHTGRSKPDTGNAVSLVPPDLTENPAGLGA 120  
 122 VEEPAAMWVGDSPVGGSELLGDDDAIYGNORSKESLGBAGI QKGSAMAATTTAIFTTLN 181  
 121 VEEPAAMWVGDSPVGGSELLGDDDAIYGNORSKESLGBAGI QKGSAMAATTTAIFTTLN 180  
 182 EPRKETRGWAKSRORROWKRRARDDQDGGISHPQWPKHSLKHYKKSPREESNQ 241  
 181 EPRKETRGWAKSRORROWKRRARDDQDGGISHPQWPKHSLKHYKKSPREESNQ 240  
 242 NGEGSYRBAETFSQVGLPLTFSGRRERLLRPEVLAETPREAFTVEAMVKEGQNN 301  
 241 NGEGSYRBAETFSQVGLPLTFSGRRERLLRPEVLAETPREAFTVEAMVKEGQNN 300  
 302 PALLAGVFNDCSHTSVSGKMAIGIRSGDKGKRDARFFSLCTDRVKKATLISHSRYP 361  
 301 PALLAGVFNDCSHTSVSGKMAIGIRSGDKGKRDARFFSLCTDRVKKATLISHSRYP 360  
 362 GYTHVAATYDGRHMAIYVDTGVASLLDQSGPLNSPFMACSRLLLGSSSEBGNHFRG 421  
 361 GYTHVAATYDGRHMAIYVDTGVASLLDQSGPLNSPFMACSRLLLGSSSEBGNHFRG 420  
 422 HLGTLVWSTALPOSHFQHSOHSSEBBAATDVLTATFEPVNTENVPFRDEKPRLEV 481  
 421 HLGTLVWSTALPOSHFQHSOHSSEBBAATDVLTATFEPVNTENVPFRDEKPRLEV 480  
 482 OGFEPEBELISPLQPLCGQTVCDNVELISQYNGYPLRGEKVI RYQVANI CDDEGLNPI 541  
 481 OGFEPEBELISPLQPLCGQTVCDNVELISQYNGYPLRGEKVI RYQVANI CDDEGLNPI 540  
 542 VSEBQIRLOHBAINEASRYNISWQLSVHOVHNSTLHRYVYL VNGEPSKI GNDHCDECE 601  
 541 VSEBQIRLOHBAINEASRYNISWQLSVHOVHNSTLHRYVYL VNGEPSKI GNDHCDECE 600  
 602 HPLTGVGDGDCRLQGR CYSMNRDGLCHVECNMNLNFDGDDCCDPVAVRKTCEPDPS 661  
 601 HPLTGVGDGDCRLQGR CYSMNRDGLCHVECNMNLNFDGDDCCDPVAVRKTCEPDPS 660  
 662 PKRAYSVKELKALQINSTHPLNTIYFASVREDLA GAATWPDKA VTHAGSIVLSPAY 721  
 661 PKRAYSVKELKALQINSTHPLNTIYFASVREDLA GAATWPDKA VTHAGSIVLSPAY 720  
 722 YGNRGHTDMIHVGHVGLGLYHVPKGVSEBSCHDPCKEIVPSNETGDL CADTAPPPKSE 781  
 721 YGNRGHTDMIHVGHVGLGLYHVPKGVSEBSCHDPCKEIVPSNETGDL CADTAPPPKSE 780  
 782 LCEBEBTSTDCGTRPRGAPFTNYMSYTDNCTDNTPQOVARMHCYLDLVYQOMTESR 841  
 781 LCEBEBTSTDCGTRPRGAPFTNYMSYTDNCTDNTPQOVARMHCYLDLVYQOMTESR 840  
 842 KPTPIPIPPNVIGQTNKSLTIHMLPISGVYDRASGLGCACTEDGTFRQYVHTASRR 901  
 841 KPTPIPIPPNVIGQTNKSLTIHMLPISGVYDRASGLGCACTEDGTFRQYVHTASRR 900

902 VCDSSGYWTEBEAVGPPVDVQCEPSLOAMSPEVHL YHNMNTVPCPTGCSLELFGHPV 961  
 901 VCDSSGYWTEBEAVGPPVDVQCEPSLOAMSPEVHL YHNMNTVPCPTGCSLELFGHPV 960  
 962 QADTLTLMWTSFPMESQVLPFTLEILLNKESVHLDPLOTFCDIPLTIKLVHNGKVSQV 1021  
 961 QADTLTLMWTSFPMESQVLPFTLEILLNKESVHLDPLOTFCDIPLTIKLVHNGKVSQV 1020  
 1022 VYTFDERIEIDALITSOHPSPLCSGCRPVRYOVLADPPFASGLPVVYTHSHKFTDVEV 1081  
 1021 VYTFDERIEIDALITSOHPSPLCSGCRPVRYOVLADPPFASGLPVVYTHSHKFTDVEV 1080  
 1082 TPQOMTQYOVLAAGELGASPLNHHIGAPYCGDKVSERLGEECDGDLVSGDCSK 1141  
 1081 TPQOMTQYOVLAAGELGASPLNHHIGAPYCGDKVSERLGEECDGDLVSGDCSK 1140  
 1142 VCELEGFNCVGEPSLCYVWYEGDICEPBRKTSIYDCGIYTPKGYLDQWATAYSHED 1201  
 1141 VCELEGFNCVGEPSLCYVWYEGDICEPBRKTSIYDCGIYTPKGYLDQWATAYSHED 1200  
 1202 KKKCPVSLVTGEPHSLICTSYHEDLPNHRPLTGMPCVASENETODDRSEQPEGLKED 1261  
 1201 KKKCPVSLVTGEPHSLICTSYHEDLPNHRPLTGMPCVASENETODDRSEQPEGLKED 1260  
 1262 EYWLKVCYFNRPGARAIFFLTIDGLVPGHQOPTVLTITDVRGNSHSLGYGLSCQHN 1321  
 1261 EYWLKVCYFNRPGARAIFFLTIDGLVPGHQOPTVLTITDVRGNSHSLGYGLSCQHN 1320  
 1322 PLIINTHHONVLFHHTTSVLANFSPRAGISVALRTSRIGLSA PSNCTISDEGQNH 1381  
 1321 PLIINTHHONVLFHHTTSVLANFSPRAGISVALRTSRIGLSA PSNCTISDEGQNH 1380  
 1382 GQSCIRPCGKQDCPSLLDHDVNVCTSIGFLMKCAITTCRGALQASGOYIRPMQ 1441  
 1381 GQSCIRPCGKQDCPSLLDHDVNVCTSIGFLMKCAITTCRGALQASGOYIRPMQ 1440  
 1442 KETILLTCSGGHNDQVNSCLPVDGVPDPSLVYANFSCSEGTFLKRSISCVPAKLOQ 1501  
 1441 KETILLTCSGGHNDQVNSCLPVDGVPDPSLVYANFSCSEGTFLKRSISCVPAKLOQ 1500  
 1502 LSPWLTCLBEDGLMSLPBYVCKLECAPIITLNNALLPHCLDNDHVGITCKYECKRGY 1561  
 1501 LSPWLTCLBEDGLMSLPBYVCKLECAPIITLNNALLPHCLDNDHVGITCKYECKRGY 1560  
 1562 VASBAGKYNRLKIKIOCLEGGIWEQSGCIPVCEBPVFEFGMYECTGFSLDSQCVLN 1621  
 1561 VASBAGKYNRLKIKIOCLEGGIWEQSGCIPVCEBPVFEFGMYECTGFSLDSQCVLN 1620  
 1622 CNOERKPLILCTKEGLWTOEFKLCENLQGECPPEPSELNVEYKCEQGYIGAVCSPLC 1681  
 1621 CNOERKPLILCTKEGLWTOEFKLCENLQGECPPEPSELNVEYKCEQGYIGAVCSPLC 1680  
 1682 VIPSPDPVWLPENITADLTLEHMEBPVQOSIYCTGRORHNPVLYVHCISQCEPQADGW 1741  
 1681 VIPSPDPVWLPENITADLTLEHMEBPVQOSIYCTGRORHNPVLYVHCISQCEPQADGW 1740  
 1742 CDTINNRAVCHYDGDCCSSTLSKVIIPFADCDLDECTCGRPAEENQ 1791  
 1741 CDTINNRAVCHYDGDCCSSTLSKVIIPFADCDLDECTCGRPAEENQ 1790

RESULT 3  
 Q9H4C9 PRELIMINARY; PRT; 1624 AA. 1022  
 Q9H4C9; 1022  
 AC 01-MAR-2001 (TRENDArel. 16, Last sequence update)  
 DT 01-MAR-2001 (TRENDArel. 16, Last sequence update)  
 DT 01-MAR-2003 (TRENDArel. 23, Last annotation update)  
 DE pregnancy-associated plasma protein-B.  
 GN PAPPE.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

